SAMPLE AVERAGE APPROXIMATION OF RISK-AVERSE STOCHASTIC PROGRAMS

A Thesis
Presented to
The Academic Faculty

by

Wei Wang

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2007

SAMPLE AVERAGE APPROXIMATION OF RISK-AVERSE STOCHASTIC PROGRAMS

Approved by:

Dr. Shabbir Ahmed, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Alexander Shapiro
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Shijie Deng
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Minqiang Li
College of Management
Georgia Institute of Technology

Dr. Martin Savelsbergh
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: August 7, 2007

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor Prof. Shabbir Ahmed for his care and fine guidance through the years. I thank him for introducing me to the fascinating field of stochastic programming. From beginning to end, I find his exuberant creativity and in-depth scholarly knowledge inspiring. I feel fortunate to have such a wonderful mentor who taught me in many aspects: cultivating new ideas, writing shell scripts, improving my presentation skills, etc.

Secondly, I am grateful to Prof. Alexander Shapiro for his many insightful suggestions on the projects I worked on. He always remains accessible to me. It is indeed a great experience to learn from one of the leaders of the field. Also I must express my gratitude to the rest of my thesis committee, Prof. Shijie Deng, Prof. Minqiang Li, and Prof. Martin Savelsbergh for their comments and valuable advice on my dissertation.

Thirdly, I thank Prof. R. Gary Parker, who brought me into this program and encouraged me to put all my effort into it. I am grateful to all the professors who taught me in the past four years. The courses I learned will be my life-time treasure.

Fourthly, I would express my appreciation to my friends here at Gatech and my friends since college. They are always there listening to me, sharing emotions, and offering help.

Finally, I thank my husband, my parents, and my sisters for their endless love and support.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

LIST OF TABLES vi

LIST OF FIGURES vii

SUMMARY viii

I INTRODUCTION 1

1.1 Traditional Stochastic Programming and Sample Average Approximation 1

1.2 Risk-Averse Stochastic Programs 4

1.3 Mean-MAD Models and Stochastic Min-Max Programs 9

1.4 CVaR and Expected Value Constrained Programs 10

1.5 Thesis Outline 13

II SAA FOR STOCHASTIC MIN-MAX PROGRAMS 15

2.1 Stochastic Min-Max Programs 15

2.1.1 Exponential Convergence Rates 16

2.1.2 Bounding Strategy and SAA Scheme 23

2.2 Mean-MAD Stochastic Programs 25

2.2.1 Reformulation as Stochastic Min-Max Programs 26

2.2.2 Convergence Results 29

2.2.3 SAA Scheme 33

2.3 Newsvendor Problem 34

III SAA FOR EXPECTED VALUE CONSTRAINED PROGRAMS 38

3.1 Expected Value Constrained Programs 38

3.1.1 Exponential Convergence Rates 39

3.1.2 Bounding Strategy and SAA Scheme 42

3.2 CVaR Constrained Programs 43

IV PORTFOLIO OPTIMIZATION 48
4.1 Introduction 48
4.2 Problem Description 51
4.3 Data Collection and Processing 52
4.4 Result Analysis 54
4.4.1 Simple Monte Carlo and Latin Hypercube Sampling Schemes 54
4.4.2 Continuous and Discrete Distributions 54
4.4.3 MAD and Variance Models 58
4.4.4 Mean-MASD, Mean-QDEV and Mean-CVaR Modes 61
4.4.5 Chance Constrained and CVaR Constrained Models 66
V SUPPLY CHAIN NETWORK DESIGN UNDER UNCERTAINTY . . . 71
5.1 Introduction 71
5.2 Problem Description and Deterministic Model 73
5.2.1 Variables 74
5.2.2 Constraints 75
5.2.3 Objective Function 77
5.3 Supply Chain Network Design under Uncertainty 77
5.4 Computational Results 81
5.4.1 Comparison of Three Models 82
5.4.2 Effects of λ 83
5.4.3 Effects of σ_d and ρ_d 83
5.4.4 Effects of p_d 83
VI CONCLUSIONS 91
REFERENCES 93

LIST OF TABLES

1.1 The SAA scheme for traditional stochastic programs 5

2.1 The SAA scheme for stochastic min-max programs 26

2.2 The SAA scheme for mean-MASD stochastic programs 34

2.3 Effects of λ (Newsvendor, mean-MASD) 36

2.4 Effects of N (Newsvendor, mean-MASD) 36

3.1 The SAA scheme for expected value constrained programs 44

4.1 SMC (Portfolio, mean-MASD, $\lambda = 0.5$, daily lognormal, APP1) . . . 55

4.2 LHS (Portfolio, mean-MASD, $\lambda = 0.5$, daily lognormal, APP1) 55

4.3 Performance of MAD and variance models (Portfolio, monthly normal, APP1) 62

4.4 Computation time of mean-risk models (Portfolio, λ=0.5, daily lognormal, APP1) 65

4.5 Effects of α (Portfolio, $q = 10$, monthly, APP1) 68

4.6 Effects of q (Portfolio, $\alpha = 95\%$, monthly, APP1) 69

5.1 Variable definition 75

5.2 Notation for supply chain network design problem 82

5.3 Comparison of three models (Supply Chain) 84

5.4 Effects of λ (Supply Chain, mean-MASD) 85

5.5 Effects of σ_d (Supply Chain, mean-MASD) 86

5.6 Effects of ρ_d (Supply Chain, mean-MASD) 87

5.7 Effects of p_d (Supply Chain, mean-MASD) 88

LIST OF FIGURES

2.1 Effects of N (corresponding to Table 2.4) 37

4.1 Efficient frontier (Portfolio, mean-MASD, APP1) 56

4.2 Portfolio composition (Portfolio, mean-MASD, APP1) 57

4.3 Three-year profits (mean-MASD, $\lambda = 0.5$) 59

4.4 Sharpe ratios (Portfolio, mean-MASD) 60

4.5 Computation time (Portfolio, monthly normal, APP1) 63

4.6 Portfolio composition (Portfolio, monthly normal, APP1) 64

4.7 Three-year profits (Portfolio, daily lognormal, APP1) 65

4.8 Effects of α and q(Portfolio, monthly, APP1) 70

5.1 Diagram of a supply chain network 74

5.2 UB and LB as functions of λ, p_d, σ_d, and ρ_d (Supply Chain) 89

5.3 $\Pr\{\text{infeas}\}_{0.95}$ as function of λ, p_d, σ_d, and ρ_d (Supply Chain) 90

SUMMARY

Sample average approximation (SAA) is a well-known solution methodology for traditional stochastic programs which are risk neutral in the sense that they consider optimization of expectation functionals. In this thesis we establish sample average approximation methods for two classes of non-traditional stochastic programs. The first class is that of stochastic min-max programs, i.e., min-max problems with expected value objectives, and the second class is that of expected value constrained stochastic programs. We specialize these SAA methods for risk-averse stochastic problems with a bi-criteria objective involving mean and mean absolute deviation, and those with constraints on conditional value-at-risk. For the proposed SAA methods, we prove that the results of the SAA problem converge exponentially fast to their counterparts for the true problem as the sample size increases. We also propose implementation schemes which return not only candidate solutions but also statistical upper and lower bound estimates on the optimal value of the true problem. We apply the proposed methods to solve portfolio selection and supply chain network design problems. Our computational results reflect good performance of the proposed SAA schemes. We also investigate the effect of various types of risk-averse stochastic programming models in controlling risk in these problems.

CHAPTER I

INTRODUCTION

1.1 Traditional Stochastic Programming and Sample Average Approximation

Optimization models derived from real world problems almost invariably involve uncertainty. Here we give three examples.

Example 1.1 Newsvendor Problem. Suppose that a newsvendor has to decide an order quantity x of next day's newspapers before observing demand d. Each newspaper costs the vendor c, can be sold at l and has a salvage value of s, where $0 \leq s < c < l$. Thus the net cost of one newspaper is $(c-l)x + (l-s)(x-d)_+$, where $[a]_+$ denotes the maximum of a and 0. According to historical data, the vendor knows the probability distribution of the demand d. To minimize the expected net cost, he formulates the following model

$$\min_{x \geq 0} \quad \mathbb{E}[(c-l)x + (l-s)(x-d)_+]. \tag{1.1}$$

Example 1.2 Portfolio Selection Problem. Consider an investor who invests a certain amount of money in a number of different assets. For each asset $k = 1, \cdots, K$, the return r_k is random. If we denote the proportion of the total funds invested in the k-th asset by x_k, this portfolio $x = (x_1, \cdots, x_K)$ has a return $r^\top x = r_1 x_1 + \cdots + r_K x_K$, which is also random. Suppose we know the probability distribution of the return vector r. We can calculate the expected return $\mathbb{E}[r^\top x]$ of this portfolio. If this investor is risk neutral and wishes to maximize the expected return, the problem can

be formulated as

$$\begin{aligned} \max \quad & \mathbb{E}[r^\top x] \\ \text{s.t.} \quad & x_1 + \cdots + x_K = 1 \\ & x_1, \cdots, x_K \geq 0. \end{aligned} \tag{1.2}$$

Note that the above problem has the trivial solution of investing only in the asset with the highest expected return.

Example 1.3 Supply Chain Network Design under Uncertainty. An operations research engineer needs to design a supply chain network for a US company that supplies cardboard packages to breweries and soft-drink manufacturers. The network consists of plants, distribution centers, customers, and transportation links from plants to distribution centers, and from distribution centers to customers. Customers have demands on cardboard packages of multiple specifications. The engineer needs to decide the locations of the plants and distribution centers from a series of candidate sites and their capacities, and furthermore, decide the material flows on transportation links. The demands from various customers are subject to market fluctuations, and in the best case, one only knows the probability distribution of the demands deduced from the historical data. If the goal is to minimize the expected cost of building and operating this supply chain, the engineer considers the following mathematical program,

$$\min_{x \in X} \quad c^\top x + \mathbb{E}[Q(x, \omega)] \tag{1.3}$$

and

$$Q(x, \omega) = \min_{y \in Y(x,\omega)} d^\top y$$

where X represents the set of feasible decisions on locations and capacities of plants and distribution centers; $c^\top x$ represents the building cost of the supply chain; ω denotes the random demands; $Y(x, \omega)$ stands for the set of feasible decisions on

material flows, depending on x and ω; and $d^\top y$ denotes the operating cost of the supply chain. We shall give a detailed explanation of this model in Chapter 5.

Formulations (1.1)-(1.3) fall into the category of traditional stochastic programming (SP), which considers optimization of an expectation functional. These problems can be generally formulated as

$$\min_{x \in X} \quad \mathbb{E}[F(x, \omega)] \tag{1.4}$$

where $X \subseteq \mathbb{R}^n$ is a nonempty set of feasible decisions, ω is a random vector having probability distribution P and support Ω, and $F : X \times \Omega \mapsto \mathbb{R}$ is a cost function such that $\mathbb{E}[F(x, \omega)]$ is well defined at all $x \in X$. Throughout this thesis, all expectations and probabilistic statements are made with respect to P unless specified otherwise. See [34] for the arguments and algorithms for various classes of stochastic programs of the form (1.4).

In many cases, exact evaluation of the expected value $\mathbb{E}[F(x, \omega)]$ in (1.4) for a given decision x is either impossible or prohibitively expensive. Suppose the random vector ω consists of L independent elements, each of which has M possible realizations, then the total number of scenarios or possible realizations of ω is M^L, which can easily grow to be astronomically large. Therefore one has to resort to approximation methods to make progress. One of the powerful methods is sample average approximation. This method and its variants are also known as stochastic counterpart methods, sample-path methods, simulated likelihood methods, etc. in the stochastic programming and statistics literature. A comprehensive review of the method can be found in [16, 34].

The idea of SAA is based on Monte Carlo simulation. An independent and identically distributed (*iid*) sample $\omega_1, \cdots, \omega_N$ of N realizations of ω is generated, and consequently the expected value function $\mathbb{E}[F(x, \omega)]$ is approximated by the sample average function $N^{-1} \sum_{n=1}^{N} F(x, \omega_n)$. The resulting sample average approximation of

stochastic program (1.4),

$$\min_{x \in X} \quad \frac{1}{N} \sum_{n=1}^{N} F(x, \omega_n), \tag{1.5}$$

is then solved by some appropriate deterministic optimization procedure. Kleywegt et al. [12] showed that under mild conditions, as the sample size increases, a solution to (1.5) converges to a solution to (1.4) with probability approaching unity exponentially fast (see also [1, 34]). This suggests that a SAA problem with a modest sample size provides a fairly good approximate solution to the original problem. In addition to the approximate optimal solutions, the SAA method also provides statistically valid upper and lower bounds on the true optimal value of (1.4). This is made possible by the following principles. The SAA value of the objective function at any feasible solution constitutes a statistical upper bound. By solving several SAA problems and taking the average of the objective values, we can obtain a statistical lower bound. This can be seen by recalling

$$\min_{x' \in X} \frac{1}{N} \sum_{n=1}^{N} F(x', \omega_n) \ \leq \ \frac{1}{N} \sum_{n=1}^{N} F(x, \omega_n), \forall x \in X,$$

by taking expectation, we have

$$\mathbb{E}\left[\min_{x' \in X} \frac{1}{N} \sum_{n=1}^{N} F(x', \omega_n)\right] \ \leq \ \mathbb{E}\left[\frac{1}{N} \sum_{n=1}^{N} F(x, \omega_n)\right] = \mathbb{E}[F(x, \omega)], \forall x \in X$$

and it follows that

$$\mathbb{E}\left[\min_{x \in X} \frac{1}{N} \sum_{n=1}^{N} F(x, \omega_n)\right] \ \leq \ \min_{x \in X} \mathbb{E}[F(x, \omega)].$$

Table 1.1 outlines the SAA scheme designed in [12]. The first two steps calculate a statistical lower and upper bound on the true objective value according to the principles given above. The last step computes an optimality gap estimate from the upper and lower bound.

1.2 Risk-Averse Stochastic Programs

Traditional stochastic programs (1.4) have an expected value objective and deterministic constraints. They also assume that the underlying probability distributions of

Table 1.1: The SAA scheme for traditional stochastic programs

Step 1:	Lower bound estimation. Generate M_l independent samples each of size N_l, i.e., $(\omega_1^m, \cdots, \omega_{N_l}^m)$ for $m = 1, \cdots, M_l$. For each sample solve the corresponding SAA problem $$\min_{x \in X} {N_l}^{-1} \sum_{n=1}^{N_l} F(x, \omega_n^m)$$ and let $\hat{v}_{N_l}^m$ and $\hat{x}_{N_l}^m$ be the corresponding optimal objective value and an optimal solution, respectively. Then one statistical lower bound and its sample variance can be estimated by $$\bar{l} = \frac{1}{M_l} \sum_{m=1}^{M_l} \hat{v}_{N_l}^m$$ $$S_{\bar{l}}^2 := \frac{1}{(M_l - 1)M_l} \sum_{m=1}^{M_l} (\hat{v}_{N_l}^m - \bar{l})^2.$$
Step 2:	Upper bound estimation. Choose a feasible solution $\bar{x} \in X$ of the true problem, for example, use one of the computed solutions $\hat{x}_{N_l}^m$. Generate an *iid* sample $(\omega_1, \cdots, \omega_{N_u})$ independent of the sample used to obtain $\bar{x}$. An estimate to the true objective function value $\mathbb{E}[F(\bar{x}, \omega)]$, which is a statistical upper bound, can be computed as $$\bar{u} = \frac{1}{N_u} \sum_{n=1}^{N_u} F(\bar{x}, \omega_n)$$ and the corresponding variance can be estimated as $$S_{\bar{u}}^2 := \frac{1}{(N_u - 1)N_u} \sum_{n=1}^{N_u} (F(\bar{x}, \omega_n) - \bar{u})^2.$$
Step 3:	Optimality gap estimation. Compute an estimate of the optimality gap of the solution $\bar{x}$ using the lower bound estimate and the objective function value estimate from Steps 1 and 2, respectively: $$\text{gap} = \bar{u} - \bar{l}.$$ The estimated variance of the above gap estimate is then given by $$S_{\text{gap}}^2 = S_{\bar{u}}^2 + S_{\bar{l}}^2.$$

the random parameters are known. These are not realistic assumptions for problems encountered in many practical situations. For a large set of problems, directly applying (1.4) would raise big concerns, and quite often one finds the performance based on the optimal solution of (1.4) is unsatisfactory. Take the portfolio selection problem as the simplest example. A portfolio with maximal expected return, obtained via solving (1.2), may have considerable possibility of serious loss. So rarely are people willing to adopt such a portfolio. This is in accordance with the fact that most people tend to be risk-averse rather than risk neutral. Another concern is related to the uncertainty of the probability distribution of random vectors. In practical applications, the probability distribution is almost never known exactly. It is then questionable whether the optimal solution from (1.4) remains good if the realized distribution deviates from the one used for computation.

Two classes of non-traditional stochastic programs have been formulated explicitly addressing the issue of risk. The first class is characterized by their risk objectives

$$\min_{x \in X} \rho[F(x, \omega)] \tag{1.6}$$

and the second class by their risk constraints

$$\min_{x \in X} \{g(x) : \rho[F(x, \omega)] \leq q\} \tag{1.7}$$

where $\rho[\cdot]$ is a risk measure, mathematically defined as a mapping assigning a real number to a measurable function from Ω to $\mathbb{R}$; $q \in \mathbb{R}$; $g(x) = \mathbb{E}[F(x, \omega)]$ or $g(\cdot)$ is some other real-valued objective function.

A well-known family of risk measures take the mean-risk form, i.e.,

$$\rho[\cdot] = \mathbb{E}[\cdot] + \lambda \mathbb{D}[\cdot], \tag{1.8}$$

where λ is a nonnegative weight parameter, and $\mathbb{D}[\cdot]$ is some dispersion statistics. In this case, (1.6) is a mean-risk model. Stochastic programs (1.6) or (1.7) involving the mean-risk-type risk measures try to reach a compromise between minimizing the cost on average and in the meanwhile reducing its variability. Note that when λ is zero, the induced problem (1.6) is of traditional SP form, and now and then we may also call $\mathbb{D}[\cdot]$ as risk measure, which can be understood as λ equals infinity in (1.8).

Let Z denote a measurable function, $\alpha \in [0, 1]$ and $\kappa_\alpha[Z]$ denotes the α-quantile of the distribution of Z. Some examples of risk measures are given below:

- mean-variance: $\mathbb{E}[Z] + \lambda \mathbb{E}[Z - \mathbb{E}Z]^2$,
- mean-mean absolute deviation (mean-MAD): $\mathbb{E}[Z] + \lambda \mathbb{E}|Z - \mathbb{E}Z|$,
- value-at-risk (VaR): $\mathrm{VaR}_\alpha[Z] := \min\{\gamma : \Pr\{Z \leq \gamma\} \geq \alpha\}$,
- mean-quantile deviation (mean-QDEV): $\mathbb{E}[Z] + \lambda \mathbb{E}[\alpha(Z - \kappa_\alpha[Z])_+ + (1 - \alpha)(\kappa_\alpha[Z] - Z)_+]$.

Note that α-quantile $\kappa_\alpha[Z]$ is any value such that $\Pr\{Z < \kappa_\alpha[Z]\} \leq \alpha \leq \Pr\{Z \leq \kappa_\alpha[Z]\}$ and the set of α-quantiles forms a closed interval, whereas $\mathrm{VaR}_\alpha[Z]$ is the smallest α-quantile from this interval.

Together with the development of the variety of risk measures are the considerable efforts devoted to developing criteria to compare the performance of risk measures. In general, we can classify the criteria into two groups. One group discusses whether a risk measure is rational. For example, a rational risk measure must satisfy the subadditivity condition, i.e., for any measurable functions Z_1 and Z_2, the risk of $Z_1 + Z_2$ is never greater than the risk of Z_1 plus the risk of Z_2, and be consistent with the rules of stochastic dominance (see [27] or Chapter 4 for details). The other group discusses whether the induced stochastic program for some risk measure is tractable. One important concept in this group is convexity-preserving. A risk measure $\rho[\cdot]$ preserves convexity if for any $F(\cdot, \omega)$ convex almost everywhere $\omega \in \Omega$, the composite function $\rho[F(\cdot, \omega)]$ is also convex. This is crucial to make (1.6) or (1.7) convex programs.

The classical Markowitz model uses mean-variance as risk measure [17, 18]. The advantage of this measure is its computation convenience in cases where $F(\cdot, \cdot)$ is bilinear, for example, $F(x, \omega) = \omega^\top x$ in Example 1.2. However, in the viewpoint of risk measurement, mean-variance has many drawbacks. First, mean and variance are measured in different units. Second, variance is a symmetric statistics and it penalizes gains and losses equally. Third, mean-variance does not preserve the convexity of cost function [2]. Fourth, mean-variance is not consistent with second-order stochastic dominance which formalizes risk-averse preferences [25]. Fifth, the variance is inappropriate to describe the risk of lower probability events, as for example the default risk.

Recently, Artzner et al.[3] introduced the very useful notion of coherent risk measure, which has become an important tool in risk management. According to [3], a real-valued function $\rho[\cdot]$, defined on a linear space $\mathcal{Z}$ of measurable functions on an

appropriate sample space, is "coherent" if it fulfills the following four criteria:

- monotonicity: If $Z_1, Z_2 \in \mathcal{Z}$ and $Z_2 \geq Z_1$, then $\rho[Z_2] \geq \rho[Z_1]$;
- subadditivity: $\rho[Z_1 + Z_2] \leq \rho[Z_1] + \rho[Z_2]$ for all $Z_1, Z_2 \in \mathcal{Z}$;
- translation invariance: If $a \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\rho[Z + a] = \rho[Z] + a$; and
- positive homogeneity: If $a > 0$ and $Z \in \mathcal{Z}$, then $\rho[aZ] = a\rho[Z]$.

By this definition, a coherent risk measure must be consistent with first-order stochastic dominance and be convexity preserving. Among the examples we gave earlier, mean-MAD (when $\lambda \in [0, 0.5]$) and mean-QDEV (when $\lambda\alpha \geq 0$ and $\lambda(1-\alpha) \in [0, 1]$) are coherent, whereas mean-variance and VaR are not [35].

Another class of non-traditional stochastic programming explicitly takes in account the uncertainty about the probability distributions. A family $\mathcal{P}$ of plausible probability distributions is considered and the optimization is carried out over all $P \in \mathcal{P}$. This leads to the min-max stochastic program

$$\min_{x \in X} \max_{P \in \mathcal{P}} \mathbb{E}_P[F(x, \omega)]. \tag{1.9}$$

A brief account of previous work on (1.9) can be found in [11]. Some recent publications build connections between the two classes of problems (1.6) and (1.9) using the concept of coherent risk measure (see [42] and references therein). It is proved that under some mild conditions, $\rho[\cdot]$ is a coherent risk measure if and only if there exists a convex set $\mathcal{P}$ of probability density functions such that

$$\rho[Z] = \max_{P \in \mathcal{P}} \mathbb{E}_P[Z], \quad \forall Z \in \mathcal{Z},$$

where $\mathcal{Z}$ denotes certain linear space of measurable functions.

In the next two sections, we shall focus on two special kinds of stochastic programs, (1.6) with mean-MAD risk measure and (1.7) with CVaR risk measure (explained later). Both problems have very nice properties and need the development of new sample average approximation methods.

1.3 Mean-MAD Models and Stochastic Min-Max Programs

The mean-MAD model with respect to cost function $F(x, \omega)$ is formulated as

$$\min_{x \in X} \ \mathbb{E}[F(x, \omega)] + \lambda \delta[F(x, \omega)] \tag{1.10}$$

where $\delta[Z] := \mathbb{E}|Z - \mathbb{E}Z|$ is the mean absolute deviation from the mean (MAD) of measurable function Z. A statistics closely related to MAD is the mean absolute semi-deviation from the mean (MASD) $\delta_+[Z] := \mathbb{E}[Z - \mathbb{E}Z]_+$. MASD seems a more reasonable risk measure than MAD since it considers only cost exceeding expectation as risk. However, owing to the equality $\delta[Z] = 2\delta_+[Z]$ as well as the existence of the weight factor, results from a mean-MASD model with factor 2λ are the same as those from a mean-MAD model with factor λ. Hence mean-MAD models and mean-MASD models are equivalent.

Mean-MAD is an attractive and promising risk measure in the area of financial engineering. This is partly due to its nice properties, for example, it is coherent and consistent with second-order stochastic dominance when $\lambda \in [0, 0.5]$ [35]. A more important reason is MAD's similarity with variance: the former is the mean "absolute" deviation from the mean, while the latter is the mean "squared" deviation from the mean. Given the shortcomings of mean-variance mentioned earlier, many works have attempted to replace variance with MAD. For example, MAD function has been introduced to economic theories, where it plays a similar role as variance in modern portfolio theory [13]. In particular, these authors showed that all capital asset pricing model-type relations for the mean-variance model also hold for the MAD model (a portfolio model with MAD as objective and constrained by mean).

While a broad class of mean-risk models such as mean-QDEV can be converted to traditional stochastic programs and solved by the existing methods, this is not the case for the mean-MAD model due to the double layers of expectation operators

in MAD. See [37] for the argument that the deterministic equivalent of the mean-MAD model does not possess the dual-block angular structure that is omnipresent in traditional two-stage stochastic programs.

Later on, we shall show that the mean-MAD model (1.10) can be reformulated as

$$\min_{x\in X, t\in\mathbb{R}} \max_{\alpha\in[0,1]} \mathbb{E}\left[F(x,\omega) + 2\lambda\alpha(F(x,\omega)-t)_+ + 2\lambda(1-\alpha)(t-F(x,\omega))_+\right].$$

This motivates us to investigate a general class of problems called **stochastic min-max programs**,

$$\min_{x\in X} \max_{y\in Y} \mathbb{E}[H(x,y,\omega)] \tag{1.11}$$

where $X \subseteq \mathbb{R}^{k_1}$ and $Y \subseteq \mathbb{R}^{k_2}$ are nonempty sets of feasible decisions, ω is a random vector with distribution P and support Ω, and $H : X \times Y \times \Omega \mapsto \mathbb{R}$ is well behaved so that its expectation is well defined for all $(x, y) \in (X, Y)$.

Note that (1.11) is related to but distinct from the **min-max stochastic programs** (1.9) discussed earlier in this section. When the set $\mathcal{P}$ of probability distributions in (1.9) takes certain special structures, for example, "Dirac structure" [38] and "band-structure" [39], or under some regularity conditions [38], min-max stochastic program (1.9) can be converted to stochastic min-max problem (1.11). Program (1.11) has not been studied extensively in literature. Shapiro [41] analyzed the asymptotics of the sample average approximation of the optimal value of (1.11), however this work did not discuss quality of solutions derived from finite samples.

1.4 CVaR and Expected Value Constrained Programs

In this section, we discuss conditional value-at-risk (CVaR) and general expected value constrained programs. CVaR is a concept derived from VaR, which is a popular measure of risk and the current standard in finance industry. In spite of its popularity, VaR has undesirable characteristics such as a lack of subadditivity and convexity, and it is coherent only when it is based on the standard deviation of normal distributions [31]. CVaR avoids these drawbacks.

We give the formal definition of CVaR following [31]. Let Z be a real-valued random variable representing loss.

- VaR is the α-percentile of loss distribution, i.e.,

$$\mathrm{VaR}_\alpha[Z] := \min\{\gamma : \ \Pr\{Z \le \gamma\} \ge \alpha\}.$$

- CVaR^+ is the expected losses strictly exceeding VaR, i.e.,

$$\mathrm{CVaR}^+_\alpha[Z] := \mathbb{E}[Z|Z > \mathrm{VaR}_\alpha[Z]].$$

- CVaR is a weighted average of VaR and CVaR^+,

$$\mathrm{CVaR}_\alpha[Z] := \theta \mathrm{VaR}_\alpha[Z] + (1-\theta)\mathrm{CVaR}^+_\alpha[Z]$$

  where

$$\theta = \frac{\Pr\{Z \le \mathrm{VaR}_\alpha[Z]\} - \alpha}{1-\alpha}.$$

Note that if Z has a continuous distribution, $\theta = 0$ and thus CVaR is equivalent to CVaR^+, denoting the expected loss exceeding VaR. It has been shown that CVaR is coherent and consistent with second-order stochastic dominance for any $\alpha \in [0, 1]$ [28, 35].

Despite the tedious definition, there is a simple formula for calculating CVaR,

$$\mathrm{CVaR}_\alpha[Z] = \min_{t \in \mathbb{R}} \left\{ t + \frac{1}{1-\alpha}\mathbb{E}[Z-t]_+ \right\}$$

with $\mathrm{VaR}_\alpha[Z]$ being the minimizer to the right-hand-side problem. This formula was shown in [31]. With this formula, problems involving CVaR can be transformed to simpler mathematical programs. For example, the mean-CVaR model

$$\min_{x \in X} \ \mathbb{E}[F(x,\omega)] + \lambda \mathrm{CVaR}_\alpha[F(x,\omega)]$$

is equivalent to

$$\min_{x \in X, t \in \mathbb{R}} \ \left\{ \mathbb{E}[F(x,\omega) + \lambda t + \frac{\lambda}{1-\alpha}\mathbb{E}[F(x,\omega) - t]_+ \right\},$$

that is,

$$\min_{x\in X, t\in\mathbb{R}} \ \mathbb{E}\left[F(x,\omega) + \lambda t + \frac{\lambda}{1-\alpha}[F(x,\omega)-t]_+\right].$$

So unlike the mean-MAD model, the mean-CVaR model can be reformulated as a traditional SP problem. A second example is the CVaR constrained problem

$$\min_{x\in X} \ \{g(x) : \mathrm{CVaR}_\alpha[F(x,\omega)] \le q\},$$

which is equivalent to

$$\min_{x\in X, t\in\mathbb{R}} \ \left\{g(x) : \mathbb{E}\left[t + \frac{1}{1-\alpha}[F(x,\omega)-t]_+\right] \le q\right\}. \tag{1.12}$$

Note that according to CVaR's definition, (1.12) is a conservative approximation of the well-known chance constrained program [21, 30] of form

$$\min_{x\in X} \{g(x) : \ \Pr\{F(x,\omega) \le q\} \ge \alpha\}, \tag{1.13}$$

i.e., any x feasible to (1.12) must be feasible to (1.13). Since usually it is difficult to solve the (non-convex) chance constrained program, solution of the CVaR constrained program can shed some light on this problem.

Problem (1.12) is a special case of the general expected value constrained programs of form

$$\min_{x\in X} \ \left\{g(x) : \mathbb{E}[\tilde{F}(x,\omega)] \le q\right\} \tag{1.14}$$

where $\tilde{F}(x,\omega)$ denotes some cost function and is such that its expectation is well defined at all $x \in X$. One characteristics of this class of problems is the existence of an expected value in the constraints.

O'Brien [24] studied (1.14) in the case where the random vector has a finite support. He proposed solution techniques including reformulating the problem as one with dual angular structure and modifying the existing Benders decomposition. Kuhn [15] and Atlason, Epelman and Henderson [4] considered the case where the support of ω is infinite. In [15], the author proposed bounding approximation schemes for multistage stochastic programs with expected value constraints. Those schemes require

that the function $\tilde{F}(x, \omega)$ is convex in both x and ω. In [4], the authors formulated a call center staffing problem as an integer expected value constrained problem and used simulation together with cutting plane method to solve it. They also analyzed the convergence rate with respect to the sample size of simulation. Their results however require that x is integer-valued.

1.5 Thesis Outline

Given their importance in risk-averse stochastic programming, it is desirable to develop solution techniques for stochastic min-max programs (1.11) and expected value constrained programs (1.14) when exactly evaluation of the expected value is impossible. This thesis is to establish sample average approximation methods for these two types of problems and test their performance via some computational studies. We believe a complete SAA method should contain two components. The first component is establishing the exponential rate at which the SAA results, both optimal solutions and objective values, converge to their true counterparts. The second component is an implementation scheme which provides solutions and evaluates their quality via upper and lower bounding. We address both these components.

The outline of this thesis is as follows. In Chapter 2, we propose a sample average approximation method for general stochastic min-max programs (1.11). This method is rather general and applicable to problems with any type of decision variables: integer, continuous or mixed-integer. We first derive its convergence rate. Then we design the SAA scheme which provides solutions and evaluates their quality via upper and lower bounding. After that, we tailor this method to mean-MAD models. The SAA scheme of mean-MAD models is slightly simpler than that of the general stochastic min-max case. Finally, we solve a newsvendor problem and compare the results from SAA with the true analytical results. Following the same course as above, Chapter 3 establishes a sample average approximation method for general expected value

constrained programs (1.14), and further specializes it to CVaR constrained problems. The next two chapters are devoted to computational studies where we apply the proposed SAA methods. In Chapter 4, we study a portfolio selection problem. We try various types of risk-averse optimization models (mean-risk models and risk constrained models) and several probability distributions (discrete and continuous distributions). In Chapter 5, we investigate a supply chain network design problem, in which demands and production costs are continuous random variables. We compare the quality of solutions from different models: deterministic, traditional SP, and mean-MASD. Finally, we conclude the thesis in Chapter 6.

CHAPTER II

SAA FOR STOCHASTIC MIN-MAX PROGRAMS

In this chapter we propose a sample average approximation method for general stochastic min-max programs and specialize it for mean-MAD models. In Section 2.1, we describe the method, derive its convergence rate, and propose a bounding strategy to make the method practical. In Section 2.2, we first show the equivalence of mean-MAD models with a special class of stochastic min-max problems, and then apply the results of Section 2.1 to the latter formulation. In Section 2.3, we report computational results on a mean-MAD model of the newsvendor problem.

2.1 Stochastic Min-Max Programs

In this section, we propose a SAA method for general stochastic min-max programs (2.1) of the following form

$$\min_{x\in X}\max_{y\in Y}\{f(x,y) := \mathbb{E}[F(x,y,\omega)]\} \tag{2.1}$$

where $X \subseteq \mathbb{R}^{k_1}$ and $Y \subseteq \mathbb{R}^{k_2}$ are nonempty sets of feasible decisions, ω is a random vector with support Ω and distribution P, and $F : X \times Y \times \Omega \mapsto \mathbb{R}$ is some function such that the expectation $\mathbb{E}[F(x,y,\omega)]$ is well defined for all $(x,y) \in (X,Y)$.

Given a random sample $\omega_1, \cdots, \omega_N$ of size N of the random vector ω, the following problem

$$\min_{x\in X}\max_{y\in Y}\left\{f_N(x,y) := \frac{1}{N}\sum_{n=1}^{N} F(x,y,\omega_n)\right\} \tag{2.2}$$

is called the sample average approximation problem for the stochastic program (2.1). Denote by v^* and (x^*, y^*) the optimal value and an optimal solution of the true problem (2.1) and by $\hat{v}_N$ and $(\hat{x}_N, \hat{y}_N)$ their counterparts of the SAA problem (2.2). We assume there is an efficient optimization algorithm for solving the (deterministic)

SAA problem for a reasonable sample size. Several crucial issues related to the SAA approach need to be addressed. (1) Whether $\hat{v}_N$ and $(\hat{x}_N, \hat{y}_N)$ converge to their true counterparts v^* and (x^*, y^*) as the sample size N is increased? (2) If so, can we analyze the convergence rate, and thereby estimate the required sample size to obtain a true optimal solution with certain confidence? (3) Can we provide any information regarding the quality of a candidate solution $(\bar{x}, \bar{y})$ for a given sample size? These questions are addressed in the subsequent subsections.

2.1.1 Exponential Convergence Rates

Our goal in this subsection is to answer the first two questions "whether converge" and "convergence rate." These questions have been successfully addressed for traditional stochastic programs. It has been shown that under mild regularity conditions, as the sample size increases, the optimal objective value and an optimal solution to the SAA problem converge with probability one (w.p.1) to their true counterparts, and moreover this convergence can be achieved at an exponential rate [1, 12, 34].

Here we directly derive "convergence rate" for stochastic min-max problems, which implies a positive answer to "whether converge." The main technique to derive the convergence rate is Large Deviations Theory, as employed by existing SAA methods for traditional stochastic programs. Let us briefly recap this important theory (see [12, 44] for detailed discussion). Consider an *iid* sequence $Z_1, \cdots, Z_N$ of replications of a real-valued random variable Z. Let $\mu := \mathbb{E}[Z]$ be its mean, which is finite, and $\hat{Z}_N = N^{-1} \sum_{n=1}^{N} Z_n$ be the corresponding sample average. Then for any real number $a > \mu$,

$$\Pr\{\hat{Z}_N \geq a\} \leq e^{-NI(a)},$$

where $I(u) := \sup_{s \in \mathbb{R}} \{su - \log M(s)\}$, $u \in \mathbb{R}$, is the large deviation (LD) rate function and $M(s) := \mathbb{E}[e^{sZ}]$ is the moment generating function (MGF) of Z. Similarly for

$a < \mu$,

$$\Pr\{\hat{Z}_N \leq a\} \leq e^{-NI(a)}.$$

Furthermore, suppose that the moment generating function $M(s)$ is finite valued in a neighborhood of $s = 0$. Then by Taylor's expansion,

$$I(a) = \frac{(a-\mu)^2}{2\sigma^2} + o(|a-\mu|^2),$$

where $\sigma^2 = \mathrm{Var}[Z]$. This implies $I(a) > 0$.

Before using Large Deviations Theory to establish convergence results for the SAA method for stochastic min-max programs, we state some required assumptions.

(A1) Both $X \subset \mathbb{R}^{k_1}$ and $Y \subset \mathbb{R}^{k_2}$ are nonempty compact sets.

(A2) The expected value function $f(x, y)$ is well defined, i.e., for any $(x, y) \in (X, Y)$, the function $F(x, y, \cdot)$ is measurable and $\mathbb{E}|F(x, y, \omega)| < +\infty$.

(A3) For any $(x, y) \in (X, Y)$, the MGF $M_{x,y}(\cdot)$ of $F(x, y, \omega) - f(x, y)$ is finite in a neighborhood of zero.

(A4) The function $F(\cdot, \cdot, \omega)$ is Lipschitz continuous almost everywhere, i.e., there exists an integrable function $\phi : \Omega \to \mathbb{R}_+$ such that for any $x_1, x_2 \in X$ and $y_1, y_2 \in Y$,

$$|F(x_1, y_1, \omega) - F(x_2, y_2, \omega)| \leq \phi(\omega) \left(\|x_1 - x_2\| + \|y_1 - y_2\|\right).$$

Then $\Phi := \mathbb{E}[\phi(\omega)]$ is finite.

(A5) The MGF $M_\phi(\cdot)$ of $\phi(\omega)$ is finite in a neighborhood of zero.

We begin with min-max problems with finite feasible solutions. Then we extend it to the case with infinite feasible solutions. Besides v^* and $\hat{v}_N$ defined before, we also define $\mathcal{S}^\epsilon$ to be the set of feasible solutions of the true problem with objective values in $[v^* - \epsilon, v^* + \epsilon]$ and $\hat{\mathcal{S}}_N$ the set of optimal solutions of the SAA problem. We show exponential convergence rates for both $\hat{v}_N$ to v^* and $\hat{\mathcal{S}}_N$ to $\mathcal{S}^\epsilon$.

Proposition 2.1. *Suppose (A1)-(A3) hold and $|X| \cdot |Y|$ is finite. Define*

$$\sigma^2 = \max_{x \in X, y \in Y} \mathrm{Var}[F(x, y, \omega) - f(x, y)].$$

Given $\epsilon > 0$ and $0 < \beta < 1$, the following is true:

(i) convergence of objective values:

$$\Pr\{|\hat{v}_N - v^*| < \epsilon\} \geq 1 - 2|X||Y|e^{-\frac{N\epsilon^2}{2\sigma^2}};$$

(ii) convergence of optimal solutions:

$$\Pr\{\hat{\mathcal{S}}_N \subset \mathcal{S}^{2\epsilon}\} \geq 1 - 4|X||Y|e^{-\frac{N\epsilon^2}{2\sigma^2}};$$

(iii) estimate of the sample size for $\Pr\{|\hat{v}_N - v^*| < \epsilon\} \geq 1 - \beta$ to hold:

$$N \geq \frac{2\sigma^2}{\epsilon^2} \log\left(\frac{2|X||Y|}{\beta}\right).$$

Proof. First, note that

$$\Pr\{|\hat{v}_N - v^*| < \epsilon\} \geq \Pr\{|f_N(x, y) - f(x, y)| < \epsilon, \ \forall (x, y) \in (X, Y)\} \tag{2.3}$$

is implied by

$$\begin{aligned}
& |f_N(x, y) - f(x, y)| < \epsilon, \ \forall (x, y) \in (X, Y) \\
\Leftrightarrow \quad & f(x, y) - \epsilon < f_N(x, y) < f(x, y) + \epsilon, \ \forall (x, y) \in (X, Y) \\
\Rightarrow \quad & \max_{y \in Y} f(x, y) - \epsilon < \max_{y \in Y} f_N(x, y) < \max_{y \in Y} f(x, y) + \epsilon, \ \forall x \in X \\
\Rightarrow \quad & \min_{x \in X} \max_{y \in Y} f(x, y) - \epsilon < \min_{x \in X} \max_{y \in Y} f_N(x, y) < \min_{x \in X} \max_{y \in Y} f(x, y) + \epsilon \\
\Leftrightarrow \quad & v^* - \epsilon < \hat{v}_N < v^* + \epsilon \\
\Leftrightarrow \quad & |\hat{v}_N - v^*| < \epsilon.
\end{aligned}$$

The right-hand-side term in (2.3) can be computed as

$$\begin{aligned}
& \Pr\{|f_N(x,y) - f(x,y)| < \epsilon,\ \forall (x,y) \in (X,Y)\} \\
= \ & 1 - \Pr\{\exists (x,y) \in (X,Y) \text{ s.t. } |f_N(x,y) - f(x,y)| \geq \epsilon\} \\
\geq \ & 1 - \sum_{x \in X, y \in Y} [\Pr\{f_N(x,y) - f(x,y) \geq \epsilon\} + \Pr\{f_N(x,y) - f(x,y) \leq -\epsilon\}] \\
\geq \ & 1 - \sum_{x \in X, y \in Y} \left[e^{-NI_{x,y}(\epsilon)} + e^{-NI_{x,y}(-\epsilon)}\right] \\
\geq \ & 1 - 2|X||Y|e^{-Na(\epsilon)},
\end{aligned}$$

where

$$a(\epsilon) := \min_{x \in X, y \in Y} \{I_{x,y}(\epsilon), I_{x,y}(-\epsilon)\}.$$

By Assumption (A3), $\mathrm{Var}[F(x,y,\omega) - f(x,y)]$ is finite at all $(x,y) \in (X,Y)$ and

$$I_{x,y}(\epsilon) \geq \frac{\epsilon^2}{2\mathrm{Var}[F(x,y,\omega) - f(x,y)]}.$$

Consequently

$$a(\epsilon) \geq \frac{\epsilon^2}{2\sigma^2} > 0.$$

Thus (i) holds.

To show (ii), we make use of the following relationships:

$$\begin{aligned}
& |f(x,y) - \hat{v}_N| < \epsilon \ \forall (x,y) \in (X,Y) \text{ and } |\hat{v}_N - v^*| < \epsilon \\
\Rightarrow \ & |f(x,y) - \hat{v}_N| < \epsilon \ \forall (x,y) \in \hat{\mathcal{S}}_N \text{ and } |\hat{v}_N - v^*| < \epsilon \\
\Leftrightarrow \ & |f(x,y) - f_N(x,y)| < \epsilon \text{ and } |f_N(x,y) - v^*| < \epsilon,\ \forall (x,y) \in \hat{\mathcal{S}}_N \\
\Rightarrow \ & |f(x,y) - f_N(x,y)| + |f_N(x,y) - v^*| < 2\epsilon \ \forall (x,y) \in \hat{\mathcal{S}}_N \\
\Rightarrow \ & |f(x,y) - v^*| < 2\epsilon \ \forall (x,y) \in \hat{\mathcal{S}}_N, \text{ i.e., } \hat{\mathcal{S}}_N \subset \mathcal{S}^{2\epsilon}.
\end{aligned}$$

Then

$$\begin{aligned}
& \Pr\{\hat{\mathcal{S}}_N \subset \mathcal{S}^{2\epsilon}\} \\
\geq\ & \Pr\{|f(x,y) - \hat{v}_N| < \epsilon \ \ \forall (x,y) \in (X,Y) \text{ and } |\hat{v}_N - v^*| < \epsilon\} \\
\geq\ & \Pr\{|f(x,y) - \hat{v}_N| < \epsilon \ \ \forall (x,y) \in (X,Y)\} + \Pr\{|\hat{v}_N - v^*| < \epsilon\} - 1 \\
\geq\ & 2\left(1 - 2|X||Y|e^{-Na(\epsilon)}\right) - 1 \\
=\ & 1 - 4|X||Y|e^{-Na(\epsilon)} \\
\geq\ & 1 - 4|X||Y|e^{-\frac{N\epsilon^2}{2\sigma^2}}.
\end{aligned}$$

So (ii) is true.

To get $\Pr\{|\hat{v}_N - v^*| < \epsilon\} \geq 1 - \beta$, it is sufficient to set $2|X||Y|e^{-\frac{N\epsilon^2}{2\sigma^2}} \leq \beta$, which gives (iii).

□

Now we consider the case where $|X| \cdot |Y|$ is infinite. Without loss of generality, we assume both $|X|$ and $|Y|$ are infinite. To deal with the infiniteness and obtain similar results as in Proposition 2.1, we use the idea of discretization as well as the assumption of Lipschitz continuity, as in existing SAA methods [1, 40]. Given $\nu > 0$, build a finite subset X_ν of X such that for any $x \in X$ there exists $x' \in X_\nu$ satisfying $\|x - x'\| \leq \nu$. Denote by D_1 the diameter of the set X, i.e., $D_1 = \max_{x_1, x_2 \in X} \|x_1 - x_2\|$. Then such set X_ν can be constructed with $|X_\nu| \leq (D_1/\nu)^{k_1}$, where k_1 is the dimension of X. In the same way, we construct Y_ν with $|Y_\nu| \leq (D_2/\nu)^{k_2}$. Note although we can choose different ν for X and Y, to get concise expressions, we take the same ν. Now fix $(x,y) \in (X,Y)$. Then there exists $(x',y') \in (X_\nu, Y_\nu)$ such that $\|x - x'\| \leq \nu$ and $\|y - y'\| \leq \nu$. By Assumption (A4), we have

$$|F(x,y,\omega) - F(x',y',\omega)| \leq 2\nu\phi(\omega),$$

which implies

$$|f(x,y) - f(x',y')| \leq 2\nu\Phi \text{ and } |f_N(x,y) - f_N(x',y')| \leq 2\nu\Phi_N,$$

where $\Phi_N = N^{-1}\sum_{n=1}^{N}\phi(\omega_n)$. Consequently, by triangle inequality we have

$$\begin{aligned} & |f_N(x,y) - f(x,y)| \\ \leq\ & |f_N(x,y) - f_N(x',y')| + |f_N(x',y') - f(x',y')| + |f(x',y') - f(x,y)| \\ \leq\ & |f_N(x',y') - f(x',y')| + 2\nu(\Phi + \Phi_N). \end{aligned} \tag{2.4}$$

Proposition 2.2. *Suppose (A1)-(A5) hold, and both $|X|$ and $|Y|$ are infinite. Define*

$$\nu := (8\Phi/\epsilon + 2)^{-1}.$$

and

$$\sigma^2 = \max_{x\in X, y\in Y}\{\mathrm{Var}[\phi(\omega)], \mathrm{Var}[F(x,y,\omega) - f(x,y)]\}.$$

Given $\epsilon > 0$ and $0 < \beta < 1$, the following inequalities are true:

(i) convergence of objective values:

$$\Pr\{|\hat{v}_N - v^*| < \epsilon\} \geq 1 - \left(1 + \frac{2D_1^{k_1}D_2^{k_2}}{\nu^{k_1+k_2}}\right)e^{-\frac{N\epsilon^2}{8\sigma^2}};$$

(ii) convergence of optimal solutions:

$$\Pr\{\hat{\mathcal{S}}_N \subset \mathcal{S}^{2\epsilon}\} \geq 1 - 2\left(1 + \frac{2D_1^{k_1}D_2^{k_2}}{\nu^{k_1+k_2}}\right)e^{-\frac{N\epsilon^2}{8\sigma^2}};$$

(iii) estimate of the sample size for $\Pr\{|\hat{v}_N - v^*| < \epsilon\} \geq 1 - \beta$ to hold:

$$N \geq \frac{8\sigma^2}{\epsilon^2}\log\left[\frac{1}{\beta}\left(1 + \frac{2D_1^{k_1}D_2^{k_2}}{\nu^{k_1+k_2}}\right)\right].$$

Proof. According to (2.4), for any $(x,y) \in (X,Y)$ satisfying $|f_N(x,y) - f(x,y)| \geq \epsilon$, there exists $(x',y') \in (X_\nu, Y_\nu)$ such that $|f_N(x',y') - f(x',y')| \geq \epsilon - 2\nu(\Phi + \Phi_N)$. We

do the bounding for $\Pr\{|f_N(x,y) - f(x,y)| < \epsilon \ \forall (x,y) \in (X,Y)\}$ as follows:

$$\begin{aligned}
& \Pr\{|f_N(x,y) - f(x,y)| < \epsilon, \ \forall (x,y) \in (X,Y)\} \\
= \ & 1 - \Pr\{\exists (x,y) \in (X,Y) \text{ s.t. } |f_N(x,y) - f(x,y)| \geq \epsilon\} \\
\geq \ & 1 - \Pr\{\exists (x,y) \in (X_\nu, Y_\nu) \text{ s.t. } |f_N(x,y) - f(x,y)| \geq \epsilon - 2\nu(\Phi + \Phi_N)\} \\
\geq \ & 1 - \Pr\{\Phi_N \geq \Phi + \epsilon/2\} - \Pr\{\exists (x,y) \in (X_\nu, Y_\nu) \text{ s.t. } |f_N(x,y) - f(x,y)| \geq \epsilon/2\} \\
\geq \ & 1 - e^{-NI_\phi(\Phi+\epsilon/2)} - \sum_{x \in X_\nu, y \in Y_\nu} \left[e^{-NI_{x,y}(\epsilon/2)} + e^{-NI_{x,y}(-\epsilon/2)} \right] \\
\geq \ & 1 - (1 + 2|X_\nu||Y_\nu|) e^{-Nb(\epsilon)} \\
\geq \ & 1 - \left[1 + 2 \left(\frac{D_1}{\nu} \right)^{k_1} \left(\frac{D_2}{\nu} \right)^{k_2} \right] e^{-Nb(\epsilon)}
\end{aligned}$$

where

$$b(\epsilon) := \min_{x \in X_\nu, y \in Y_\nu} \{ I_\phi(\Phi + \epsilon/2), I_{x,y}(\epsilon/2), I_{x,y}(-\epsilon/2) \}.$$

By Assumptions (A3) and (A5),

$$I_\phi(\Phi + \epsilon/2) \geq \frac{\epsilon^2}{8\mathrm{Var}[\phi(\omega)]},$$

$$I_{x,y}(\epsilon/2) \geq \frac{\epsilon^2}{8\mathrm{Var}[F(x,y,\omega) - f(x,y)]},$$

and $\mathrm{Var}[\phi(\omega)]$ and $\mathrm{Var}[F(x,y,\omega) - f(x,y)]$ for all $(x,y) \in (X,Y)$ are finite valued. Furthermore, (A4) ensures that $\mathrm{Var}[F(x,y,\omega) - f(x,y)]$ is continuous on the compact sets X and Y, which implies that σ^2 must be finite. Therefore

$$b(\epsilon) \geq \frac{\epsilon^2}{8\sigma^2} > 0.$$

Now all results can be obtained in a way similar to Proposition 2.1.

□

The proofs of the two propositions imply a positive answer to "whether converge." A second way to show "converge" is by utilizing the relationship of uniform convergence, similar to that for the SAA method for traditional stochastic programs. Here we give a proof sketch. By the law of large numbers, we know $f_N(x,y)$ converges to

$f(x, y)$ pointwise. If we can further show $f_N(x, y)$ converges to $f(x, y)$ uniformly on (X, Y) w.p.1, then for any $x \in X$, the maximum of $f_N(x, y)$ over $y \in Y$ converges to that of $f(x, y)$ uniformly w.p.1, and so $\hat{v}_N$ to v^*. It is easy to see the uniform convergence of $f_N(x, y)$ to $f(x, y)$ in cases where (X, Y) only contains a finite number of points (by pointwise convergence). In cases where (X, Y) contains an infinite number of points, the Lipschitz continuity ensures the uniform convergence (see Proposition 7 on page 363 in [34]).

2.1.2 Bounding Strategy and SAA Scheme

In the previous subsection we obtained the convergence results for the SAA method for stochastic min-max programs. These provide theoretical justification for the proposed method. Yet they have two drawbacks: (1) it is difficult to calculate the constants σ^2 in the exponential terms, due to difficulty in computing the LD rate function; (2) even if we can calculate those constants, in general the estimated sample size is too conservative to be practical. So in practice, we use relatively small size samples and get some candidate solution. How should we validate this solution? This is Question (3) raised at the beginning of this section. Here we describe techniques to estimate the optimality gap of a candidate solution obtained from the SAA problem with finite sample size.

The structure of stochastic min-max programs indicates that by fixing any $x \in X$, the inner maximization problem provides an upper bound on the optimal value of the original min-max problem, i.e.,

$$u^*(x) \geq v^* \quad \forall x \in X, \tag{2.5}$$

where

$$u^*(x) := \max_{y \in Y} \mathbb{E}[F(x, y, \omega)]; \tag{2.6}$$

similarly, fixing any $y \in Y$, the outer minimization problem would give a lower bound

on the optimal value of the original min-max problem, i.e.,

$$l^*(y) \leq v^* \quad \forall y \in Y, \tag{2.7}$$

where

$$l^*(y) := \min_{x \in X} \mathbb{E}[F(x, y, \omega)]. \tag{2.8}$$

The equality in (2.5) or (2.7) can be attained when $x \in X^*$ or $y \in Y^*$, with X^* and Y^* denoting the set of optimal x-values and the set of optimal y-values to the original min-max problems. More important, Problem (2.6) and (2.8) are traditional stochastic programs, for which we know how the existing SAA method generates statistical upper and lower bounds. Therefore to acquire a good statistical upper (lower) bound for v^*, it is sufficient to find $\tilde{x}$ ($\tilde{y}$), an estimate for some $x \in X^*$ ($y \in Y^*$), as well as a good statistical upper bound for $u^*(\tilde{x})$ (lower bound for $l^*(\tilde{y})$).

We acquire $(\tilde{x}, \tilde{y})$ by directly solving the SAA problem for the original min-max problem. Again, let $\omega_1, \cdots, \omega_N$ be a sample of size N. Solve the SAA problem

$$\min_{x \in X} \max_{y \in Y} \frac{1}{N} \sum_{n=1}^{N} F(x, y, \omega_n)$$

and let $(\tilde{x}, \tilde{y})$ be an optimal solution.

Now let us recall how to get a statistical upper bound for $u^*(\tilde{x})$. Let $\omega_1, \cdots, \omega_{N_u}$ be another sample of size N_u. Solve the SAA problem to (2.6) with $x = \tilde{x}$,

$$\hat{u}_{N_u}(\tilde{x}) := \max_{y \in Y} \left\{ f_{N_u}(\tilde{x}, y) := \frac{1}{N_u} \sum_{n=1}^{N_u} F(\tilde{x}, y, \omega_n) \right\}.$$

The following is a well-known result:

$$\begin{aligned} & \max_{y \in Y} f_{N_u}(\tilde{x}, y) \;\geq\; f_{N_u}(\tilde{x}, y') \quad \forall y' \in Y \\ \Longrightarrow \quad & \mathbb{E}\left[\max_{y \in Y} f_{N_u}(\tilde{x}, y) \right] \;\geq\; f(\tilde{x}, y') \quad \forall y' \in Y \\ \Longrightarrow \quad & \mathbb{E}\left[u_{N_u}(\tilde{x}) \right] \;\geq\; u^*(\tilde{x}) := \max_{y' \in Y} f(\tilde{x}, y'). \end{aligned}$$

That is, $\mathbb{E}[\hat{u}_{N_u}(\tilde{x})] \geq u^*(\tilde{x})$. By generating M_u independent samples of the uncertain parameters, each of size N_u, and solving the corresponding SAA problems, we obtain

M_u optimal objective values $\hat{u}^1_{N_u}(\tilde{x}), \cdots, \hat{u}^{M_u}_{N_u}(\tilde{x})$. Then

$$\bar{u} := \frac{1}{M_u} \sum_{m=1}^{M_u} \hat{u}^m_{N_u}(\tilde{x})$$

is an unbiased estimator of $\mathbb{E}[\hat{u}_{N_u}(\tilde{x})]$ and therefore is a statistical upper bound on $u^*(\tilde{x})$. Note that an estimate of variance of the above estimator can be computed as

$$S^2_{\bar{u}} := \frac{1}{M_u(M_u - 1)} \sum_{m=1}^{M_u} \left[\hat{u}^m_{N_u}(\tilde{x}) - \bar{u}\right]^2.$$

Define z_β by $\Pr\{Z \leq z_\beta\} = 1 - \beta$ where Z is a standard normal random variable. When M_u is large, a $(1-\beta)$-confidence interval for $\mathbb{E}[\hat{u}_{N_u}(\tilde{x})]$ can be calculated as

$$\left[\bar{u} - z_{\beta/2} S_{\bar{u}},\ \bar{u} + z_{\beta/2} S_{\bar{u}}\right].$$

When M_u is small, $z_{\beta/2}$ can be replaced by $t_{\beta/2, M_u - 1}$, the critical value of a t-distribution. A statistically valid lower bound for $l^*(\tilde{y})$ can be obtained in a similar way. By generating M_l samples each of size N_l and solving the SAA problem to (2.8), we get the statistical lower bound $\bar{l}$ for v^* and its sample variance $S^2_{\bar{l}}$, and in turn, a confidence interval.

The sizes of N, N_u, and M_u should be a trade-off between computational work and accuracy requirement. The larger the sample sizes, the smaller the variance estimates, but the heavier the computational work. We can adjust these sizes through some preliminary experiments. In the computational part, we also explore the possibility of reducing variances via variance reduction techniques. We summarize the SAA scheme in Table 2.1.

2.2 Mean-MAD Stochastic Programs

In the remainder of this chapter, we concentrate on mean-MAD stochastic programs. As we already analyzed, a mean-MAD model is equivalent to a mean-MASD model with doubled weight factor. Since the latter can simplify our statement, we discuss the mean-MASD models instead. We first prove the equivalence of mean-MASD

Table 2.1: The SAA scheme for stochastic min-max programs

Step 1:	Optimal $(\tilde{x}, \tilde{y})$ estimation. Generate a sample of size N, i.e., $(\omega_1, \cdots, \omega_N)$. Solve the SAA problem $\min_{x\in X} \max_{y\in Y} f_N(x, y)$ and let $(\tilde{x}, \tilde{y})$ be an optimal solution.
Step 2:	Lower bound estimation. Generate M_l independent samples each of size N_l, i.e., $(\omega_1^m, \cdots, \omega_{N_l}^m)$ for $m = 1, \cdots, M_l$. For each sample, solve the lower bounding SAA problem $\hat{l}_{N_l}^m(\tilde{y}) := \min_{x\in X} f_{N_l}(x, \tilde{y})$. Compute the lower bound estimate $\bar{l}$ and its variance $S_{\bar{l}}^2$ as follows $\bar{l} := \frac{1}{M_l} \sum_{m=1}^{M_l} \hat{l}_{N_l}^m(\tilde{y})$, $S_{\bar{l}}^2 := \frac{1}{M_l(M_l-1)} \sum_{m=1}^{M_l} \left[\hat{l}_{N_l}^m(\tilde{y}) - \bar{l}\right]^2$ and a $(1-\beta)$-confidence interval for this lower bound is $\left[\bar{l} - \gamma S_{\bar{l}},\ \bar{l} + \gamma S_{\bar{l}}\right]$, where $\gamma = t_{\beta/2, M_l-1}$ when M_l is small; otherwise $\gamma = z_{\beta/2}$.
Step 3:	Upper bound estimation. Generate M_u independent samples each of size N_u, i.e., $(\omega_1^m, \cdots, \omega_{N_u}^m)$ for $m = 1, \cdots, M_u$. For each sample, solve the upper bounding SAA problem $\hat{u}_{N_u}^m(\tilde{x}) := \max_{y\in Y} f_{N_u}(\tilde{x}, y)$. Compute the upper bound estimate $\bar{u}$ and its variance $S_{\bar{u}}^2$ $\bar{u} := \frac{1}{M_u} \sum_{m=1}^{M_u} \hat{u}_{N_u}^m(\tilde{x})$, $S_{\bar{u}}^2 := \frac{1}{M_u(M_u-1)} \sum_{m=1}^{M_u} \left[\hat{u}_{N_u}^m(\tilde{x}) - \bar{u}\right]^2$, and a $(1-\beta)$-confidence interval for this upper bound is $\left[\bar{u} - \gamma S_{\bar{u}},\ \bar{u} + \gamma S_{\bar{u}}\right]$, where $\gamma = t_{\beta/2, M_u-1}$ for small M_u and $\gamma = z_{\beta/2}$ for large M_u.

models with one certain min-max problems. Then we employ the results obtained in the previous section to this particular kind of min-max problems. We specialize the assumptions required for obtaining the convergence rates for the SAA method on this kind of stochastic min-max programs.

2.2.1 Reformulation as Stochastic Min-Max Programs

Let us restate the definition of a mean-MASD stochastic program. Let $G(x, \omega)$ be a cost function on $X \times \Omega$ and λ be a nonnegative real value. Suppose $\mathbb{E}[G(x, \omega)]$ is well defined at every $x \in X$. The mean-MASD stochastic program associated with $G(\cdot, \cdot)$ is defined as

$$\min_{x\in X} \mathbb{E}[G(x, \omega)] + \lambda \delta_+[G(x, \omega)], \tag{2.9}$$

where $\delta_+[Z] = \mathbb{E}[Z - \mathbb{E}Z]_+$ is the MASD of some random variable Z. A closely related problem is the mean-QDEV stochastic program

$$\min_{x \in X} \mathbb{E}[G(x, \omega)] + \lambda h_\alpha[G(x, \omega)], \tag{2.10}$$

where $h_\alpha[Z] := \mathbb{E}[\alpha(Z - \kappa_\alpha[Z])_+ + (1 - \alpha)(\kappa_\alpha[Z] - Z)_+]$ is the α-quantile deviation and $\kappa_\alpha[Z]$ is the α-quantile of Z. Ogryczak and Ruszczyński [27] showed that MASD is indeed the worst-case quantile deviation, i.e.,

$$\delta_+[Z] = \max_{\alpha \in [0,1]} h_\alpha[Z]. \tag{2.11}$$

A well-known formula for α-quantile deviation is

$$h_\alpha[Z] = \min_{t \in \mathbb{R}} \mathbb{E}[\alpha(Z - t)_+ + (1 - \alpha)(t - Z)_+] \tag{2.12}$$

with $\kappa_\alpha[Z]$ a minimizer to the right-hand-side problem [27]. With this formula, the mean-QDEV model (2.10) can be reformulated as

$$\min_{x \in X, t \in \mathbb{R}} \mathbb{E}[\tilde{G}_\lambda(x, t, \alpha, \omega)]$$

where

$$\tilde{G}_\lambda(x, t, \alpha, \omega) := G(x, \omega) + \lambda\alpha[G(x, \omega) - t]_+ + \lambda(1 - \alpha)[t - G(x, \omega)]_+. \tag{2.13}$$

The equalities (2.11) and (2.12) allow us to reformulate a mean-MASD model as a min-max problem.

Proposition 2.3. *Suppose $\mathbb{E}[G(x, \omega)]$ is well defined at every $x \in X$. Then for any $\lambda \geq 0$, the mean-MASD stochastic program (2.9) is equivalent to the stochastic min-max program*

$$\min_{x \in X, t \in \mathbb{R}} \max_{\alpha \in [0,1]} \mathbb{E}[\tilde{G}_\lambda(x, t, \alpha, \omega)] \tag{2.14}$$

with $\tilde{G}_\lambda(x, t, \alpha, \omega)$ defined in (2.13).

Proof. We can derive as follows:

$$\begin{aligned}
&\min_{t\in\mathbb{R}} \max_{\alpha\in[0,1]} \mathbb{E}[\tilde{G}_\lambda(x,t,\alpha,\omega)] && (2.15)\\
=\ & \mathbb{E}[G(x,\omega)] + \lambda \min_{t\in\mathbb{R}} \max_{\alpha\in[0,1]} \{\alpha\mathbb{E}[G(x,\omega)-t]_+ + (1-\alpha)\mathbb{E}[t-G(x,\omega)]_+\}\\
=\ & \mathbb{E}[G(x,\omega)] + \lambda \min_{t\in\mathbb{R}} \max \{\mathbb{E}[G(x,\omega)-t]_+,\ \mathbb{E}[t-G(x,\omega)]_+\}\\
=\ & \mathbb{E}[G(x,\omega)] + \lambda\mathbb{E}[G(x,\omega)-\mathbb{E}G(x,\omega)]_+ && (2.16)
\end{aligned}$$

where the last two equalities are due to the fact that for any random variable Z with finite mean $\mathbb{E}[Z]$, the following holds

$$\mathbb{E}[Z-\mathbb{E}Z]_+ = \mathbb{E}[\mathbb{E}Z - Z]_+.$$

Note that in (2.15), the optimal t-value is $\mathbb{E}[G(x,\omega)]$ and the optimal α-value is such that $\kappa_\alpha[G(x,\omega)] = \mathbb{E}[G(x,\omega)]$. Now taking minimization over $x \in X$ to (2.15) and (2.16) at the same time, we have

$$\min_{x\in X} \{\mathbb{E}[G(x,\omega)] + \lambda\delta_+[G(x,\omega)]\} = \min_{x\in X, t\in\mathbb{R}} \max_{\alpha\in[0,1]} \mathbb{E}[\tilde{G}_\lambda(x,t,\alpha,\omega)].$$

This completes the proof. □

The function $\tilde{G}_\lambda(x,t,\alpha,\omega)$ has many desirable properties. Let us fix $\lambda \in [0,1]$. The following concise analysis is from [41]. By defining

$$H_\lambda(\alpha,z) := z + \lambda\alpha[z]_+ + \lambda(1-\alpha)[-z]_+,$$

we have

$$\tilde{G}_\lambda(x,t,\alpha,\omega) = H_\lambda[\alpha, G(x,\omega)-t] + t.$$

Clearly $H_\lambda(\alpha,z)$ is linear in α and convex in z. Moreover, $H_\lambda(\alpha,z)$ is nondecreasing in z. So $H_\lambda(\alpha,\cdot)$ is convexity preserving. Therefore, if $G(x,\omega)$ is convex in x, then $\tilde{G}_\lambda(x,t,\alpha,\omega)$ is linear (thus concave) in α and convex in x and t, and in turn Problem (2.14) is a convex program.

2.2.2 Convergence Results

We make the following assumptions.

(B1) $X \subset \mathbb{R}^k$ is a nonempty compact set. X can be either discrete or continuous. Without loss of generality, we assume X is continuous here.

(B2) For any $x \in X$, $G(x, \cdot)$ is measurable.

(B3) There exists an integrable function $\psi : \Omega \to \mathbb{R}_+$ such that at some $\tilde{x} \in X$, $|G(\tilde{x}, \omega)| \leq \psi(\omega)$ for every $\omega \in \Omega$. Let $\Psi := \mathbb{E}[\psi(\omega)]$.

(B4) The MGF of $\psi(\omega)$, $M_\psi(s) = \mathbb{E}[e^{s\psi(\omega)}]$ is finite valued in a neighborhood of zero.

(B5) There exists an integrable function $\phi : \Omega \to \mathbb{R}_+$ such that

$$|G(x_1, \omega) - G(x_2, \omega)| \leq \phi(\omega)\|x_1 - x_2\|, \ \forall \omega \in \Omega \text{ and } x_1, x_2 \in X.$$

Let $\Phi := \mathbb{E}[\phi(\omega)]$.

(B6) The MGF of $\phi(\omega)$, $M_\phi(s) = \mathbb{E}[e^{s\phi(\omega)}]$ is finite valued in a neighborhood of zero.

We use $\tilde{G}$ to denote $\tilde{G}_\lambda(x, t, \alpha, \omega)$. To apply the convergence result of general stochastic min-max programs in Section 2.1, we need to make sure that conditions (A1)-(A5) are satisfied. Below we make the following proofs in sequence: (1) prove the finiteness of $\mathbb{E}|\tilde{G}|$ and as a byproduct, define a compact set for t; (2) verify the Lipschitz continuity of $\tilde{G}$, and (3) show the finiteness in a neighborhood of zero of the moment generating function $M_{x,t,\alpha}(\cdot)$ of $(\tilde{G} - \mathbb{E}[\tilde{G}])$.

Lemma 2.4. *Under Assumptions (B1)-(B6), the following statements are true.*

(i) Let D be the diameter of X. Then $\mathbb{E}|G(x, \omega)| \leq \Psi + \Phi D$ for all $x \in X$.

(ii) $\mathbb{E}|\tilde{G}_\lambda(x, t, \alpha, \omega)| < +\infty$ for all $x \in X$, $\alpha \in [0, 1]$ and finite t.

(iii) The domain of t in Problem (2.14) can be replaced by a bounded interval $T := [-\Phi D - \Psi, \Phi D + \Psi]$ without changing the optimization results.

Proof. Since X is compact, $D < +\infty$. By Assumption (B5),

$$|G(x,\omega) - G(\tilde{x},\omega)| \leq \phi(\omega)\|x - \tilde{x}\| \leq \phi(\omega)D, \quad \forall x \in X.$$

It follows that

$$|G(x,\omega)| \leq |G(\tilde{x},\omega)| + \phi(\omega)D \leq \psi(\omega) + \phi(\omega)D, \quad \forall x \in X, \tag{2.17}$$

where the second inequality is due to Assumption (B3). Assumptions (B4) and (B6) give the finiteness of Ψ and Φ. So (i) is true. Furthermore, inequality (2.17) also implies that for any $x \in X$, finite t and $\alpha \in [0,1]$, $\mathbb{E}[G(x,\omega) - t]_+$ and $\mathbb{E}[t - G(x,\omega)]_+$ are finite as well. Together with inequality

$$\mathbb{E}|\tilde{G}| \leq \mathbb{E}|G(x,\omega)| + \lambda\alpha\mathbb{E}[G(x,\omega) - t]_+ + \lambda(1-\alpha)\mathbb{E}[t - G(x,\omega)]_+,$$

it follows that (ii) holds.

From the discussion in the previous subsection, $t^* = \mathbb{E}[G(x^*,\omega)]$, where (x^*, t^*, α^*) denotes one optimal solution of (2.14). Then according to (i), $|t^*|$ can be bounded by $\Phi D + \Psi$. Consequently, in the min-max problem (2.14), substituting $t \in \mathbb{R}$ by $t \in T$ will not change the optimal value and solutions, while the feasible region becomes compact. This completes the proof.

□

Lemma 2.5. *Let $x_1, x_2 \in X$, $\alpha_1, \alpha_2 \in [0,1]$ and $t_1, t_2 \in \mathbb{R}$ finite. Then under Assumptions (B1)-(B6), a.e. $\omega \in \Omega$,*

$$\begin{aligned} & |\tilde{G}_\lambda(x_1,t_1,\alpha_1,\omega) - \tilde{G}_\lambda(x_2,t_2,\alpha_2,\omega)| \qquad (2.18) \\ \leq\ & (1+\lambda)\phi(\omega)\|x_1 - x_2\| + \lambda|t_1 - t_2| + \lambda[\Psi + \psi(\omega) + (\Phi + \phi(\omega))D]|\alpha_1 - \alpha_2|. \end{aligned}$$

Proof. The Lipschitz continuity of $\tilde{G}$ with respect to x modulus $(1+\lambda)\phi(\omega)$ can be shown as follows. Let G, G_1 and G_2 be short notations for $G(x,\omega)$, $G(x_1,\omega)$ and $G(x_2,\omega)$, respectively. Then

$$\begin{aligned}
& |\tilde{G}_\lambda(x_1,t,\alpha,\omega) - \tilde{G}_\lambda(x_2,t,\alpha,\omega)| \\
= \; & |G_1 - G_2 + \lambda\alpha[(G_1-t)_+ - (G_2-t)_+] + \lambda(1-\alpha)[(t-G_1)_+ - (t-G_2)_+]| \\
\leq \; & |G_1 - G_2| + \lambda\alpha|G_1 - G_2| + \lambda(1-\alpha)|G_1 - G_2| \\
\leq \; & (1+\lambda)|G_1 - G_2| \\
\leq \; & (1+\lambda)\phi(\omega)\|x_1 - x_2\|.
\end{aligned}$$

Similarly, $\tilde{G}$ is Lipschitz continuous with respect to t modulus λ since

$$\begin{aligned}
& |\tilde{G}_\lambda(x,t_1,\alpha,\omega) - \tilde{G}_\lambda(x,t_2,\alpha,\omega)| \\
\leq \; & \lambda\alpha|(G-t_1)_+ - (G-t_2)_+| + \lambda(1-\alpha)|(t_1-G)_+ - (t_2-G)_+| \\
\leq \; & \lambda|t_1 - t_2|.
\end{aligned}$$

Finally, $\tilde{G}$ is Lipschitz continuous with respect to α modulus $\lambda[\Psi + \psi(\omega) + (\Phi + \phi(\omega))D]$ because

$$\begin{aligned}
& |\tilde{G}_\lambda(x,t,\alpha_1,\omega) - \tilde{G}_\lambda(x,t,\alpha_2,\omega)| \\
= \; & \lambda\,|[G-t]_+ + [t-G]_+|\,|\alpha_1 - \alpha_2| \\
= \; & \lambda|G-t||\alpha_1 - \alpha_2| \\
\leq \; & \lambda(|G| + |t|)|\alpha_1 - \alpha_2| \\
\leq \; & \lambda[\Psi + \psi(\omega) + (\Phi + \phi(\omega))D]|\alpha_1 - \alpha_2|.
\end{aligned}$$

Hence by triangle inequality, (2.18) holds.

□

Lemma 2.6. *Let $x \in X$, $\alpha \in [0,1]$ and t be finite real-valued. Then under Assumptions (B1)-(B6), the MGF $M_{x,t,\alpha}(\cdot)$ of $(\tilde{G} - \mathbb{E}[\tilde{G}])$ is finite around zero.*

Proof. We have already shown that $\mathbb{E}|\tilde{G}| < +\infty$. So $e^{s\mathbb{E}[\tilde{G}]}$ is finite around zero. On the other hand, by Assumption (B4), both terms on the right hand side of the following equation are finite around zero,

$$\mathbb{E}[e^{sG(x,\omega)}] = \int_{\{\omega:G\leq t\}} e^{sG} dP(\omega) + \int_{\{\omega:G>t\}} e^{sG} dP(\omega);$$

consequently, both terms on the right hand side of the following equation are also finite around zero,

$$\mathbb{E}[e^{s\tilde{G}}] = e^{s\lambda(1-\alpha)t} \int_{\omega:G\leq t} e^{s[1-\lambda(1-\alpha)]G} dP(\omega) + e^{-s\lambda\alpha t} \int_{\omega:G>t} e^{s(1+\lambda\alpha)G} dP(\omega).$$

It follows that $M_{x,t,\alpha}(s)$ is finite around zero.

□

Proposition 2.7. *Suppose Assumptions (B1)-(B6) hold. Define*

$$\nu := \left[1 + 2\lambda + \lambda D + \frac{4\lambda + 4\Phi(1+\lambda) + 8\lambda(\Phi D + \Psi)}{\epsilon}\right]^{-1}$$

and

$$\sigma^2 := \max_{x\in X, t\in T, \alpha\in[0,1]} \left\{ \mathrm{Var}[\phi(\omega)], \mathrm{Var}[\psi(\omega)], \mathrm{Var}[\tilde{G}_\lambda(x,t,\alpha,\omega) - \mathbb{E}[\tilde{G}_\lambda(x,t,\alpha,\omega)]] \right\}.$$

Given $\epsilon > 0$ and $0 < \beta < 1$, we have

(i) convergence of objective values:

$$\Pr\{|\hat{v}_N - v^*| < \epsilon\} \geq 1 - 2\left[1 + 2(\Phi D + \Psi)\frac{D^k}{\nu^{k+2}}\right] e^{-\frac{N\epsilon^2}{8\sigma^2}},$$

(ii) convergence of optimal solutions:

$$\Pr\{\hat{\mathcal{S}}_N \subset \mathcal{S}^{2\epsilon}\} \geq 1 - 4\left[1 + 2(\Phi D + \Psi)\frac{D^k}{\nu^{k+2}}\right] e^{-\frac{N\epsilon^2}{8\sigma^2}};$$

(iii) estimate of the sample size for $\Pr\{|\hat{v}_N - v^*| < \epsilon\} \geq 1 - \beta$ to hold:

$$N \geq \frac{8\sigma^2}{\epsilon^2} \log\left\{\frac{2}{\beta}\left[1 + 2(\Phi D + \Psi)\frac{D^k}{\nu^{k+2}}\right]\right\}.$$

2.2.3 SAA Scheme

In this part we detail the SAA method for solving mean-MASD stochastic programs. Comparing with a general stochastic min-max program, here we have two equivalent formulations: one is a minimization problem and the other is a min-max problem. We shall take advantage of this speciality and design a simpler SAA method for mean-MASD problems.

As before, we fix α at some $\tilde{\alpha}$ and solve M_l SAA problems to the outer minimization problem of the min-max formulation (i.e., the mean-QDEV stochastic program) to get a stochastically lower bound. Let $\omega_1, \cdots, \omega_N$ be one sample of size N. $\tilde{\alpha}$ can be obtained via directly solving the SAA problem of the original mean-MASD problem (2.9)

$$\min_{x \in X} \left\{ \frac{1}{N} \sum_{n=1}^{N} G(x, \omega_n) + \frac{\lambda}{N} \sum_{n=1}^{N} \left[G(x, \omega_n) - \frac{1}{N} \sum_{j=1}^{N} G(x, \omega_j) \right]_+ \right\},$$

for example, using a decomposition-based method [2]. Let $\tilde{x}$ be one of its optimal solutions. Then

$$\tilde{\alpha} = \frac{|Q|}{N}, \text{ where } Q = \left\{ \omega_n : \ n = 1, \cdots, N, \ G(\tilde{x}, \omega_n) < \frac{1}{N} \sum_{j=1}^{N} G(\tilde{x}, \omega_j) \right\}.$$

The approach for constructing a statistical upper bound is simulation. Consider a feasible solution $\bar{x}$, for example, taking it to be $\tilde{x}$ or one of solutions from lower bounding. Generate another independent sample of size N_u and calculate the quantities

$$G'_n := G(\bar{x}, \omega_n) + \lambda \left[G(\bar{x}, \omega_n) - \frac{1}{N_u} \sum_{j=1}^{N_u} G(\bar{x}, \omega_j) \right]_+ \text{ and } \bar{u} := \frac{1}{N_u} \sum_{n=1}^{N_u} G'_n.$$

Since $\bar{u}$ is an unbiased estimator of $\mathbb{E}[G(\bar{x}, \omega) + \lambda \delta_+[G(\bar{x}, \omega)]]$, it is a statistical upper bound for the optimal value of the mean-MASD model. The SAA scheme for mean-MASD problems is given in Table 2.2.

Table 2.2: The SAA scheme for mean-MASD stochastic programs

Step 1:	Optimal $\tilde{\alpha}$ estimation. Generate a sample of size N, i.e., $(\omega_1, \cdots, \omega_N)$. Solve the following SAA problem $\min_{x\in X} \left\{ \frac{1}{N}\sum_{n=1}^{N} G(x,\omega_n) + \frac{\lambda}{N}\sum_{n=1}^{N}\left[G(x,\omega_n) - \frac{1}{N}\sum_{j=1}^{N} G(x,\omega_j)\right]_+ \right\}$ and let $\tilde{x}$ be its optimal solution and $g_N(\tilde{x}) = N^{-1}\sum_{j=1}^{N} G(\tilde{x},\omega_j)$. Then $\tilde{\alpha}$ can be computed as $\tilde{\alpha} = \frac{\lvert Q\rvert}{N},$ where $Q = \{\omega_n :\ n = 1, \cdots, N,\ G(\tilde{x},\omega_n) < g_N(\tilde{x})\}.$
Step 2:	Lower bound estimation. Generate M_l independent samples each of size N_l, i.e., $(\omega_1^m, \cdots, \omega_{N_l}^m)$ for $m = 1, \cdots, M_l$. For each sample, solve the SAA mean-QDEV problem at $\tilde{\alpha}$ $\hat{l}^m := \min_{x\in X, t\in\mathbb{R}} \frac{1}{N_l}\sum_{n=1}^{N_l} [G(x,\omega_n^m) + \lambda\tilde{\alpha}(G(x,\omega_n^m) - t)_+ + \lambda(1-\tilde{\alpha})(t - G(x,\omega_n^m))_+].$ Compute the lower bound estimate $\bar{l}$ and its variance $S_{\bar{l}}^2$ as follows $\bar{l} := \frac{1}{M_l}\sum_{m=1}^{M_l} \hat{l}^m,$ $S_{\bar{l}}^2 := \frac{1}{M_l(M_l-1)}\sum_{m=1}^{M_l}(\hat{l}^m - \bar{l})^2,$ and a $(1-\beta)$-confidence interval for this lower bound is $\left[\bar{l} - \gamma S_{\bar{l}},\ \bar{l} + \gamma S_{\bar{l}}\right],$ where $\gamma = t_{\beta/2, M_l-1}$ when M_l is small; otherwise $\gamma = z_{\beta/2}$.
Step 3:	Upper bound estimation. Choose a feasible solution $\bar{x} \in X$ of the mean-MASD problem, for example, $\tilde{x}$ or one of the solutions from Step 2. Generate an independent sample of size N_u, i.e., $(\omega_1, \cdots, \omega_{N_u})$. Let $g_{N_u}(\bar{x}) = N_u^{-1}\sum_{j=1}^{N_u} G(\bar{x},\omega_j)$. Compute the upper bound estimate $\bar{u}$ and its variance $S_{\bar{u}}^2$ as $G'_n := G(\bar{x},\omega_n) + \lambda[G(\bar{x},\omega_n) - g_{N_u}(\bar{x})]_+,$ $\bar{u} := \frac{1}{N_u}\sum_{n=1}^{N_u} G'_n,$ $S_{\bar{u}}^2 := \frac{1}{N_u(N_u-1)}\sum_{n=1}^{N_u}(G'_n - \bar{u})^2,$ and a $(1-\beta)$-confidence interval for this upper bound is $\left[\bar{u} - z_{\beta/2}S_{\bar{u}},\ \bar{u} + z_{\beta/2}S_{\bar{u}}\right],$ since M_u is usually very large.

2.3 Newsvendor Problem

In this section we computationally investigate the proposed SAA method. We consider a single period newsvendor problem without setup costs. Each newspaper costs the vendor c, can be sold at l and has a salvage value of s, where $s < c < l$. The newspaper demand ω is uniformly distributed on $[a, b]$ (we choose this simple distribution so as to compute the mean-MASD objective exactly). Before the start of the day, newsvendor needs to determine the order quantity of the newspaper, represented

by x. Thus the net cost of the newspapers is

$$G(x,\omega) = (c-l)x + (l-s)(x-\omega)_+.$$

In this case, the mean-MASD model can be solved analytically. First we derive the following terms

$$\mathbb{E}[G(x,\omega)] = (c-l)x + \frac{(l-s)(x-a)^2}{2(b-a)},$$
$$\delta_+[G(x,\omega)] = \frac{(l-s)(x-a)^2(2b-a-x)^2}{8(b-a)^3}.$$

Then for any λ, let $f_\lambda(x) := \mathbb{E}[G(x,\omega)] + \lambda\delta_+[G(x,\omega)]$. Solve the cubic equation

$$\frac{df_\lambda(x)}{dx} = 0,$$

which provides us one optimal solution x^* and the optimal value v^* of the mean-MASD model. Utilizing the fact that $\kappa_{\alpha^*}[\cdot] = \mathbb{E}[\cdot]$, we get one optimal α^* to the minimax equivalent of the mean-MASD model:

$$\alpha^* = \frac{1}{2}\left[1 + \frac{(b-x^*)^2}{2(b-a)^2}\right].$$

We also use the SAA approach to solve the same model and compare the results with the analytical solutions. In all experiments, we set $c = 10$, $l = 20$, $s = 2$, $a = 100$, and $b = 400$. The optimal objective value of the traditional stochastic programming model (i.e., $\min_{x\in X}\mathbb{E}[G(x,\omega)]$) is -1833.33 under this data set. Table 2.3 presents results for different λ, where SAA parameters are $N = 2000$, $N_l = 1000$, $M_l = 10$, and $N_u = 20000$. $\bar{x}$ is the one, among M_l optimal solutions obtained in the lower bounding step, providing the lowest $\bar{u}$-value. Table 2.4 and Figure 2.1 give results for different N at $\lambda = 0.5$, and other SAA parameters are $N_l = 2000$, $M_l = 20$, and $N_u = 100000$. Please refer to Table 2.3 for the true value and optimal solution. $\tilde{x}$ is the solution obtained in Step 1 of Table 2.2, $f(\tilde{x})$ is the true objective value at $\tilde{x}$ and $f_{N_u}(\tilde{x})$ is this value approximated by a sample of size N_u. In Figure 2.1, $\hat{v}_N$

turns to be steadily close to v^* when $N > 12000$, which is an expensive sample size. However, this does not mean we have to use such large sample sizes to obtain good solutions. In fact, the solution from $N = 500$ is good enough, which can be seen by the closeness of $\hat{v}_N$ to $\bar{l}$ and $f_{N_u}(\tilde{x})$.

Table 2.3: Effects of λ ($N = 2000$, $N_l = 1000$, $M_l = 10$ and $N_u = 20000$)

λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
α^*	0.607	0.616	0.625	0.635	0.646	0.657	0.668	0.679	0.690
x^*	261.21	255.57	249.78	243.90	237.98	232.07	226.25	220.57	215.07
$-v^*$	1790.74	1749.97	1711.10	1674.19	1639.31	1606.47	1575.65	1546.83	1519.94
$\tilde{\alpha}$	0.612	0.617	0.629	0.639	0.648	0.663	0.673	0.682	0.690
$\bar{x}$	260.71	254.93	248.58	243.48	237.68	233.59	227.07	220.87	214.28
$-\bar{u}$	1786.59	1745.80	1706.92	1669.90	1634.81	1601.65	1570.63	1541.67	1514.76
$S_{\bar{u}}$	7.15	7.29	7.36	7.48	7.53	7.65	7.59	7.51	7.36
$-\bar{l}$	1790.88	1749.86	1710.78	1673.82	1638.97	1606.13	1575.17	1546.23	1519.46
$S_{\bar{l}}$	7.33	7.05	6.79	6.49	6.18	5.91	5.69	5.50	5.37

Table 2.4: Effects of N ($\lambda = 0.5$, $N_l = 2000$, $M_l = 20$ and $N_u = 100000$)

N	100	500	1000	2000	4000	6000	8000	10000	12000
$\tilde{\alpha}$	0.640	0.636	0.643	0.640	0.644	0.646	0.645	0.643	0.645
$\tilde{x}$	222.02	237.24	229.69	229.85	235.70	235.39	237.28	237.50	238.22
$-\hat{v}_N$	1544.83	1627.15	1591.81	1579.72	1624.74	1626.30	1625.87	1622.41	1636.84
$-f(\tilde{x})$	1631.80	1639.30	1637.30	1637.38	1639.16	1639.12	1639.30	1639.31	1639.31
$-f_{N_u}(\tilde{x})$	1633.08	1641.09	1638.80	1639.39	1640.74	1640.20	1640.77	1640.52	1641.73
$S_{f_{N_u}(\tilde{x})}$	2.91	3.34	3.13	3.13	3.29	3.28	3.34	3.34	3.36
$-\bar{l}$	1640.36	1639.26	1640.66	1641.79	1641.07	1641.49	1641.12	1641.60	1636.98
$S_{\bar{l}}$	6.99	5.57	5.26	6.67	6.55	6.52	6.58	6.48	5.33

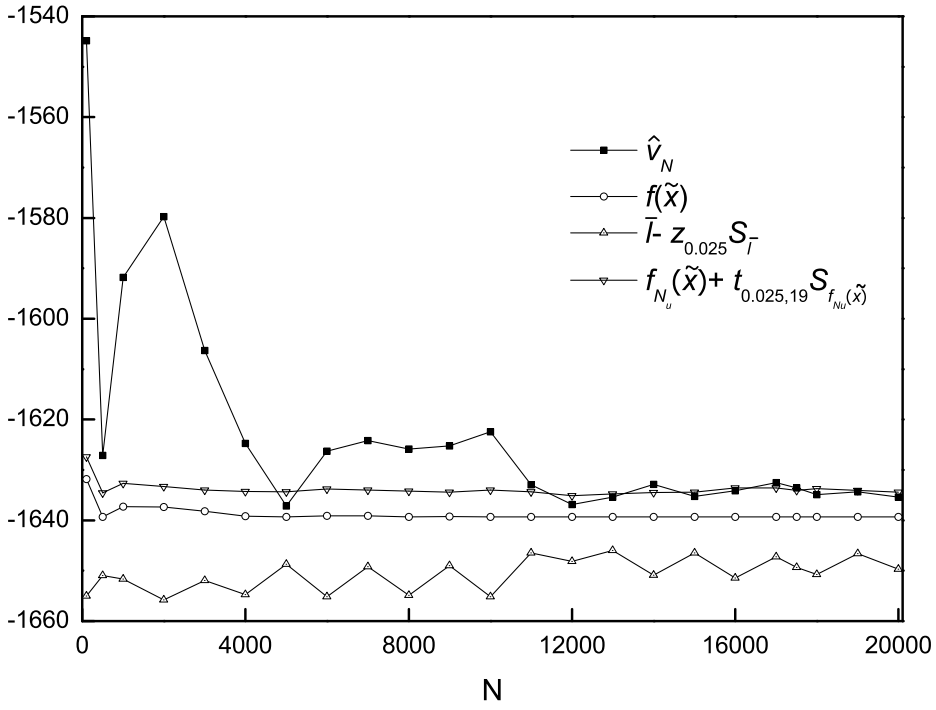


Figure 2.1: Effects of N (corresponding to Table 2.4)

CHAPTER III

SAA FOR EXPECTED VALUE CONSTRAINED PROGRAMS

This chapter is devoted to developing a sample average approximation method for expected value constrained programs. In Section 3.1, we derive the convergence rates and design the SAA scheme. In Section 3.2, we specialize the proposed method to the CVaR constrained programs.

3.1 Expected Value Constrained Programs

Consider a stochastic program defined on the following region

$$\{x \in X \text{ s.t. } g(x) := \mathbb{E}[G(x,\omega)] \leq q\}, \tag{3.1}$$

where $X \subseteq \mathbb{R}^k$ is a nonempty set of feasible decisions, ω is a random vector with support Ω and distribution P, and $G : X \times \Omega \mapsto \mathbb{R}$ is some function such that the expectation $\mathbb{E}[G(x,\omega)]$ is well defined for all $x \in X$.

Feasible region (3.1) can arise in many situations. For example, as we pointed out in Chapter 1, a CVaR constrained problem can be transformed into a program constrained on the set (3.1). A second example is any stochastic program where the goal is to minimize the risk such that the mean is below some value. A third example is a two-stage stochastic program in which the objective is to minimize the first-stage cost such that the expected second-stage cost is below some value. When applying the SAA method proposed here, we require the expected value constraint in (3.1) to be soft, i.e., after a slight adjustment to the q-value, the problem remains mathematically feasible and meaningful. In cases where the approximation is "restrictive" (the region is smaller than that of the original problem), we assume the problem is still feasible.

On the other hand, in cases where the approximation is "relaxed" (the region is bigger), we hope the solution obtained is still meaningful in the sense that it can provide some information to the real-world problem.

3.1.1 Exponential Convergence Rates

Let $\epsilon > 0$. Define $X^\epsilon := \{x \in X : g(x) \leq q + \epsilon\}$. Due to difficulty in evaluating $\mathbb{E}[G(\cdot, \omega)]$ accurately, we consider using a sample average to approximate it. Let $\omega_1, \cdots, \omega_N$ be a sample of size N of the random vector ω. Then $\mathbb{E}[G(x, \omega)]$ can be approximated by $g_N(x) := N^{-1} \sum_{n=1}^{N} G(x, \omega_n)$. Consequently we define $X_N^\epsilon := \{x \in X : g_N(x) \leq q + \epsilon\}$. Our goal is to estimate

$$\Pr\{X^{-\epsilon} \subseteq X_N^0 \subseteq X^\epsilon\},$$

that is, we want to claim a feasible solution of certain SAA problem is feasible to the true problem but not too conservative. Using Large Deviations Theory, this would give us the exponential convergence results for the SAA method of the expected value constrained stochastic programs. We state some required assumptions.

(C1) $X \subset \mathbb{R}^k$ is nonempty compact sets.

(C2) The expected value function $g(x)$ is well defined, i.e., for every $x \in X$, the function $G(x, \cdot)$ is measurable and $\mathbb{E}|G(x, \omega)| < +\infty$.

(C3) For any $x \in X$, the moment generating function $M_x(\cdot)$ of $G(x, \omega) - g(x)$ is finite in a neighborhood of zero.

(C4) For any $\omega \in \Omega$ there exists an integrable function $\phi : \Omega \to \mathbb{R}_+$ such that

$$|G(x_1, \omega) - G(x_2, \omega)| \leq \phi(\omega)\|x_1 - x_2\|, \ \forall x_1, x_2 \in X.$$

Denote $\Phi := \mathbb{E}[\phi(\omega)]$.

(C5) The MGF $M_\phi(\cdot)$ of $\phi(\omega)$ is finite in a neighborhood of zero.

Proposition 3.1. *Suppose (C1)-(C3) hold and $|X|$ is finite. Given $\epsilon > 0$, then*

$$\Pr\left\{X^{-\epsilon} \subseteq X_N^0 \subseteq X^{\epsilon}\right\} \geq 1 - 2|X|e^{-\frac{N\epsilon^2}{2\sigma^2}}$$

where $\sigma^2 := \max_{x\in X} \mathrm{Var}[G(x,\omega) - g(x)]$.

Proof.

$$\begin{aligned}
& \Pr\left\{X^{-\epsilon} \subseteq X_N^0 \subseteq X^{\epsilon}\right\} \\
= \;& 1 - \Pr\{\exists x \in X \text{ s.t. } g(x) \leq q - \epsilon \text{ and } g_N(x) > q, \\
& \qquad \text{or } \exists x \in X \text{ s.t. } g_N(x) \leq q \text{ and } g(x) > q + \epsilon\} \\
\geq \;& 1 - \Pr\{\exists x \in X \text{ s.t. } g_N(x) - g(x) > \epsilon\} - \Pr\{\exists x \in X \text{ s.t. } g_N(x) - g(x) < -\epsilon\} \\
\geq \;& 1 - \sum_{x\in X} \left[\Pr\{g_N(x) - g(x) > \epsilon\} + \Pr\{g_N(x) - g(x) < -\epsilon\}\right] \\
\geq \;& 1 - \sum_{x\in X} \left[e^{-NI_x(\epsilon)} + e^{-NI_x(-\epsilon)}\right] \\
\geq \;& 1 - 2|X|e^{-Na(\epsilon)},
\end{aligned}$$

where $a(\epsilon) := \min_{x\in X}\{I_x(\epsilon), I_x(-\epsilon)\}$. Assumption (C3) implies that $I_x(\epsilon)$ and $I_x(-\epsilon)$ are no less than $\epsilon^2/(2\mathrm{Var}[G(x,\omega) - g(x)])$ at any $x \in X$, and as a result, $a(\epsilon) \geq \epsilon^2/(2\sigma^2)$.

□

For given $\nu > 0$, build a finite subset X_ν of X such that for any $x \in X$ there exists $x' \in X_\nu$ satisfying $\|x - x'\| \leq \nu$. Denoting by D the diameter of the set X, i.e., $D = \max_{x_1,x_2\in X} \|x_1 - x_2\|$, then such set X_ν can be constructed with $|X_\nu| \leq (D/\nu)^k$. Now fix $x \in X$ and $x' \in X_\nu$ satisfying $\|x - x'\| \leq \nu$. Suppose Assumption (C4) holds. Then

$$|G(x,\omega) - G(x',\omega)| \leq \phi(\omega)\nu.$$

Consequently

$$|g(x) - g(x')| \leq \Phi\nu$$

and

$$|g_N(x) - g_N(x')| \leq \Phi_N\nu$$

where $\Phi_N = N^{-1}\sum_{n=1}^{N}\phi(\omega_n)$.

Proposition 3.2. *Suppose (C1)-(C5) hold and $|X|$ is infinite. Given $\epsilon > 0$, then*

$$\Pr\{X^{-\epsilon} \subseteq X_N^0 \subseteq X^{\epsilon}\} \geq 1 - 2\left[1 + \left(\frac{D}{\nu}\right)^k\right] e^{-\frac{N\epsilon^2}{8\sigma^2}},$$

where $\nu := (4\Phi/\epsilon + 1)^{-1}$ and $\sigma^2 := \max_{x\in X}\{\mathrm{Var}[\phi(\omega)], \mathrm{Var}[G(x,\omega) - g(x)]\}$.

Proof.

$$\begin{aligned}
& \Pr\{X^{-\epsilon} \subseteq X_N^0 \subseteq X^{\epsilon}\} \\
\geq\ & 1 - \Pr\{\exists x \in X \text{ s.t. } g_N(x) - g(x) > \epsilon\} - \Pr\{\exists x \in X \text{ s.t. } g_N(x) - g(x) < -\epsilon\} \\
\geq\ & 1 - \Pr\{\exists x \in X_\nu \text{ s.t. } g_N(x) - g(x) > \epsilon - (\Phi + \Phi_N)\nu\} \\
& \quad - \Pr\{\exists x \in X_\nu \text{ s.t. } g(x) - g_N(x) > \epsilon - (\Phi + \Phi_N)\nu\} \\
\geq\ & 1 - 2\Pr\{\Phi_N > \Phi + \epsilon/2\} - \Pr\{\exists x \in X_\nu \text{ s.t. } g_N(x) - g(x) > \epsilon/2\} \\
& \quad - \Pr\{\exists x \in X_\nu \text{ s.t. } g(x) - g_N(x) > \epsilon/2\} \\
\geq\ & 1 - 2e^{-NI_\phi(\Phi+\epsilon/2)} - \sum_{x\in X_\nu}\left[e^{-NI_x(\epsilon/2)} + e^{-NI_x(-\epsilon/2)}\right] \\
\geq\ & 1 - 2(1 + |X_\nu|)e^{-Nb(\epsilon)} \\
\geq\ & 1 - 2\left[1 + \left(\frac{D}{\nu}\right)^k\right] e^{-Nb(\epsilon)}
\end{aligned}$$

where $b(\epsilon) := \min_{x\in X_\nu}\{I_\phi(\Phi + \epsilon/2), I_x(\epsilon/2), I_x(-\epsilon/2)\}$. By Assumptions (C3) and (C5),

$$\begin{aligned}
b(\epsilon) &\geq \min_{x\in X_\nu}\left\{\frac{\epsilon^2}{8\mathrm{Var}[\phi(\omega)]}, \frac{\epsilon^2}{8\mathrm{Var}[G(x,\omega) - g(x)]}\right\} \\
&= \frac{\epsilon^2}{8\max_{x\in X_\nu}\{\mathrm{Var}[\phi(\omega)], \mathrm{Var}[G(x,\omega) - g(x)]\}} \\
&\geq \frac{\epsilon^2}{8\sigma^2}.
\end{aligned}$$

Thus the statement is true. □

3.1.2 Bounding Strategy and SAA Scheme

Now we discuss bounding strategy and design a SAA scheme for expected value constrained programs.

From Lagrange duality, we know

$$\min_{x\in X} \{f(x) : g(x) := \mathbb{E}[G(x,\omega)] \le q\} \tag{3.2}$$

$$\geq \min_{x\in X} \{f(x) + \pi[g(x) - q]\} \tag{3.3}$$

for any $\pi \geq 0$ and the equality holds when π is an optimal Lagrangian multiplier. Since (3.3) takes the form of traditional stochastic programming, we can use it to generate a lower bound for (3.2). As to upper bounding, we know any $x \in X$ satisfying $g(x) \leq q$ can provide an upper bound $f(x)$. However, for expected value constrained problems, it is difficult to tell if $x \in X$ is really feasible to the original problem, i.e., whether it satisfies $g(x) \leq q$. Therefore, we need to associate an upper bound obtained via evaluating the objective function value $f(x)$ at some x with an estimated probability that it is a true upper bound.

We obtain $\tilde{x}$ for upper bounding and $\tilde{\pi}$ for lower bounding via solving one SAA problem of the original problem. Let $\omega_1, \cdots, \omega_N$ be a sample of size N. Solve the approximate SAA problem,

$$\min_{x\in X} \{f(x) : g_N(x) := N^{-1}\sum_{n=1}^{N} G(x,\omega_n) \le \tilde{q}\} \tag{3.4}$$

and let $(\tilde{x}, \tilde{\pi})$ be the primal-dual optimizer pair. Note $\tilde{q} \leq q$. In order to solve (3.4) efficiently, N cannot be too large. Consequently it is very possible that $\tilde{x}$ from solving (3.4) with $\tilde{q} = q$ is infeasible. Therefore a smaller $\tilde{q}$-value in (3.4) can improve the chance that $\tilde{x}$ is feasible. Let $\omega_1, \cdots, \omega_{N_u}$ be a sample of size N_u, where $N_u \gg N$. Compute

$$\tilde{u} := \frac{1}{N_u}\sum_{n=1}^{N_u} G(\tilde{x}, \omega_n)$$

and

$$S_{\tilde{u}}^2 := \frac{1}{N_u(N_u-1)} \sum_{n=1}^{N_u} [G(\tilde{x}, \omega_n) - \tilde{u}]^2.$$

For N_u large enough, we know

$$\frac{g(\tilde{x}) - \tilde{u}}{S_{\tilde{u}}} \sim N(0,1).$$

Similarly as before, define z_β by $\Pr\{Z \leq z_\beta\} = 1 - \beta$, where Z is a standard normal random variable and $\beta \in [0,1]$. Let

$$z_\beta = \frac{q - \tilde{u}}{S_{\tilde{u}}}.$$

Then

$$\Pr\{g(\tilde{x}) \leq q\} = 1 - \beta.$$

So we only need to check z_β. If z_β is big enough, then we accept $\tilde{x}$. Otherwise we continue to decrease $\tilde{q}$ by some small value ξ, and resolve (3.4) to get another $\tilde{x}$ and check it. After obtaining a satisfactory $\tilde{x}$ with characteristic z_β, we can calculate $f(\tilde{x})$ and conclude that the probability that $f(\tilde{x})$ is an upper bound is β. Table 3.1 describes the SAA scheme for expected value constrained problems.

3.2 CVaR Constrained Programs

Let $\alpha \in (0,1)$ and Z be a real-valued random variable. As introduced in Chapter 1,

$$\mathrm{CVaR}_\alpha[Z] = \min_{t \in \mathbb{R}} \left\{ t + \frac{1}{1-\alpha} \mathbb{E}[Z - t]_+ \right\}, \tag{3.5}$$

and one of the minimizers equals $\mathrm{VaR}_\alpha[Z]$. In this section, we consider the following problem

$$\min_{x \in X} \ \{f(x) : \mathrm{CVaR}_\alpha[G(x,\omega)] \leq q\}, \tag{3.6}$$

where $q \in \mathbb{R}$, $f : X \to \mathbb{R}$, and $G : X \times \Omega \to \mathbb{R}$ such that $\mathrm{CVaR}_\alpha[G(x,\omega)]$ is well defined at all $x \in X$. By (3.5) and defining

$$\tilde{G}_\alpha(x,t,\omega) = t + \frac{1}{1-\alpha}[G(x,\omega) - t]_+,$$

Table 3.1: The SAA scheme for expected value constrained programs

Step 0:	Set $\tilde{z} > 0$, $\xi > 0$, $\gamma \in (0,1)$ and $\tilde{q} = q$.
Step 1:	Optimal $\tilde{\pi}$ estimation. Generate a sample of size N, i.e., $(\omega_1, \cdots, \omega_N)$ and solve the following SAA problem $\min \{f(x) : x \in X, g_N(x) \leq \tilde{q}\}$. Let $\tilde{x}$ and $\tilde{\pi}$ be its optimal solution and Lagrange multiplier, respectively.
Step 2:	Upper bound estimation. Generate another independent sample of size N_u, i.e., $(\omega_1, \cdots, \omega_{N_u})$. Compute $g_{N_u}(\tilde{x})$, $S^2_{N_u}(\tilde{x})$ and $z_\beta = \frac{q - g_{N_u}(\tilde{x})}{S_{\tilde{u}}(\tilde{x})}$. If $z_\beta < \tilde{z}$, let $\tilde{q} = \tilde{q} - \xi$ and go to step 1; otherwise get an upper bound $\tilde{u}$ and the associated possibility $\tilde{u} = f(\tilde{x})$, $\Pr\{\tilde{u} \text{ is an upper bound}\} = \beta$.
Step 3:	Lower bound estimation. Generate M_l independent samples each of size N_l, i.e., $(\omega_1^m, \cdots, \omega_{N_l}^m)$ for $m = 1, \cdots, M_l$. For each sample, solve the SAA problem $\hat{l}^m := \min_{x \in X} f(x) + \tilde{\pi} \left[g^m_{N_l}(x) - q \right]$. Compute the lower bound estimator $\tilde{l}$ and its variance $S^2_{\tilde{l}}$ as follows $\tilde{l} := \frac{1}{M_l} \sum_{m=1}^{M_l} \hat{l}^m$, $S^2_{\tilde{l}} := \frac{1}{M_l(M_l-1)} \sum_{m=1}^{M_l} (\hat{l}^m - \tilde{l})^2$. And a $(1-\gamma)$ confidence interval for $\mathbb{E}[\tilde{l}]$ is $[\tilde{l} - z_{\gamma/2} S_{\tilde{l}},\ \tilde{l} + z_{\gamma/2} S_{\tilde{l}}]$.

Problem (3.6) is equivalent to

$$\min_{x \in X, t \in \mathbb{R}} \{f(x) :\ \mathbb{E}[\tilde{G}_\alpha(x,t,\omega)] \leq q\}, \tag{3.7}$$

to which we can apply the SAA method of expected value constrained problems, and the convergence results in Section 3.1 hold provided that (C1)-(C5) are satisfied. Here we make the following assumptions.

(D1) $X \subset \mathbb{R}^k$ is a nonempty compact set. X can be either discrete or continuous. Without loss of generality, we assume X is continuous here.

(D2) For any $x \in X$, $G(x, \cdot)$ is measurable.

(D3) There exists a measurable function $\psi : \Omega \to \mathbb{R}_+$ such that $|G(x,\omega)| \leq \psi(\omega)$ for every $x \in X$ and $\omega \in \Omega$. Let $\Psi := \mathbb{E}[\psi(\omega)]$.

(D4) The MGF of $\psi(\omega)$, $M_\psi(s) = \mathbb{E}[e^{s\psi(\omega)}]$ is finite valued in a neighborhood of zero.

(D5) There exists a measurable function $\phi : \Omega \to \mathbb{R}_+$ such that

$$|G(x_1, \omega) - G(x_2, \omega)| \leq \phi(\omega)\|x_1 - x_2\|, \ \forall \omega \in \Omega \text{ and } x_1, x_2 \in X.$$

Let $\Phi := \mathbb{E}[\phi(\omega)]$.

(D6) The MGF of $\phi(\omega)$, $M_\phi(s) = \mathbb{E}[e^{s\phi(\omega)}]$ is finite valued in a neighborhood of zero.

Lemma 3.3. *Let $x_1, x_2 \in X$ and $t_1, t_2 \in \mathbb{R}$. Then under Assumptions (D1)-(D6), a.e. $\omega \in \Omega$,*

$$|\tilde{G}_\alpha(x_1, t_1, \omega) - \tilde{G}_\alpha(x_2, t_2, \omega)| \leq \frac{\phi(\omega)}{1-\alpha}\|x_1 - x_2\| + \frac{2-\alpha}{1-\alpha}|t_1 - t_2|. \tag{3.8}$$

Proof. The Lipschitz continuity of $\tilde{G}$ with respect to x can be shown as follows. Let G_1 and G_2 be short notations for $G(x_1, \omega)$ and $G(x_2, \omega)$, respectively. Then

$$\begin{aligned} |\tilde{G}_\alpha(x_1, t, \omega) - \tilde{G}_\alpha(x_2, t, \omega)| &= \frac{1}{1-\alpha}|(G(x_1, \omega) - t)_+ - (G(x_2, \omega) - t)_+| \\ &\leq \frac{1}{1-\alpha}|G(x_1, \omega) - G(x_2, \omega)| \\ &\leq \frac{\phi(\omega)}{1-\alpha}\|x_1 - x_2\|. \end{aligned}$$

Similarly, $\tilde{G}$ is Lipschitz continuous with respect to t since

$$\begin{aligned} |\tilde{G}_\alpha(x, t_1, \omega) - \tilde{G}_\alpha(x, t_2, \omega)| &= |t_1 - t_2| + \frac{1}{1-\alpha}|(G(x, \omega) - t_1)_+ - (G(x, \omega) - t_2)_+| \\ &\leq |t_1 - t_2| + \frac{1}{1-\alpha}|t_1 - t_2| \\ &= \frac{2-\alpha}{1-\alpha}|t_1 - t_2|. \end{aligned}$$

Hence by triangle inequality, inequality (3.8) holds.

□

Lemma 3.4. *Suppose Assumptions (D1)-(D4) hold. There exists a closed interval $T \subset \mathbb{R}$ such that*

$$\min_{x \in X, t \in T} \{f(x) : \ \mathbb{E}[\tilde{G}_\alpha(x, t, \omega)] \leq q\} \tag{3.9}$$

is equivalent to the problem (3.7) where t is a free variable.

Proof. Fixing $x \in X$, we know

$$\mathrm{VaR}_\alpha[G(x,\omega)] \in \mathrm{Argmin}_{t\in\mathbb{R}}\mathbb{E}[\tilde{G}_\alpha(x,t,\omega)].$$

Therefore, it is enough to bound $\mathrm{VaR}_\alpha[G(x,\omega)]$ for all $x \in X$ instead to bound t.

According to Assumption (D3),

$$-\psi(\omega) \le G(x,\omega) \le \psi(\omega),\ \forall x \in X \text{ and } \omega \in \Omega.$$

It follows that

$$\mathrm{VaR}_\alpha[-\psi(\omega)] \le \mathrm{VaR}_\alpha[G(x,\omega)] \le \mathrm{VaR}_\alpha[\psi(\omega)],\ \forall x \in X.$$

By Assumption (D4), $\mathbb{E}[\psi(\omega)] < +\infty$. Together with the fact $\psi(\omega) \ge 0$, it is true that $\mathrm{VaR}_\alpha[\psi(\omega)] < +\infty$ and $\mathrm{VaR}_\alpha[-\psi(\omega)] > -\infty$ for any $\alpha \in (0,1)$. Hence we can define

$$T := [\mathrm{VaR}_\alpha[-\psi(\omega)], \mathrm{VaR}_\alpha[\psi(\omega)]]. \tag{3.10}$$

□

Lemma 3.5. *Under Assumptions (D1)-(D4), for any $x \in X$ and $t < \infty$, the moment generating function of $M_{x,t,\alpha}(\cdot)$ of $\tilde{G}_\alpha(x,t,\omega) - \mathbb{E}[\tilde{G}_\alpha(x,t,\omega)]$ is finite around zero.*

Proof.

$$M_{x,t,\alpha}(s) := e^{-s\mathbb{E}[\tilde{G}_\alpha(x,t,\omega)]}\mathbb{E}\left[e^{s\tilde{G}_\alpha(x,t,\omega)}\right].$$

The finiteness of $\mathbb{E}[\tilde{G}_\alpha(x,t,\omega)]$ is trivial since $\mathbb{E}[G(x,\omega)-t]_+ < +\infty$, which is implied by $\mathbb{E}|G(x,\omega)| \le \mathbb{E}[\psi(\omega)] < +\infty$. So it is enough to prove the finiteness of $\mathbb{E}\left[e^{s\tilde{G}_\alpha(x,t,\omega)}\right]$, which can be split into two terms

$$e^{\frac{s\alpha t}{\alpha-1}}\int_{\omega:G(x,\omega)>t} e^{\frac{s}{1-\alpha}G(x,\omega)}dP(\omega) + e^{st}\mathrm{Pr}\{\omega : G(x,\omega) \le t\}.$$

Since

$$\int_{\omega:G(x,\omega)>t} e^{\frac{s}{1-\alpha}G(x,\omega)}dP(\omega) = \begin{cases} \mathbb{E}\left[e^{\frac{s}{1-\alpha}\psi(\omega)}\right] & s > 0, \\ \mathbb{E}\left[e^{\frac{-s}{1-\alpha}\psi(\omega)}\right] & s \le 0, \end{cases}$$

by Assumption (D4), it is finite. So $M_{x,t,\alpha}(\cdot) < +\infty$ in a neighborhood of zero.

□

Denote the feasible region of Problem (3.7) by Y^0 and that of the sample approximate problem by Y_N^0. Finally we have the following convergence result.

Proposition 3.6. *Suppose Assumptions (D1)-(D6) hold. Given $\epsilon > 0$, then it is true that*

$$\Pr\{Y^{-\epsilon} \subseteq Y_N^0 \subseteq Y^{\epsilon}\} \geq 1 - 2\left[1 + \left(\frac{D_x}{\nu}\right)^k \left(\frac{D_t}{\nu}\right)\right] e^{-\frac{N\epsilon^2}{8\sigma^2}},$$

where

- *k and D_x are the dimension and the diameter of X, respectively,*
- *D_t is the diameter of T, e.g. $T = [\mathrm{VaR}_\alpha[-\psi(\omega)], \mathrm{VaR}_\alpha[\psi(\omega)]]$,*
- *$\nu := \left\{\frac{1}{1-\alpha}\left[\frac{4(\Phi-\alpha+2)}{\epsilon} + 1\right]\right\}^{-1}$, and*
- *$\sigma^2 := \max_{x\in X, t\in T}\{\mathrm{Var}[\phi(\omega)], \mathrm{Var}[\tilde{G}_\alpha(x,t,\omega) - \mathbb{E}[\tilde{G}_\alpha(x,t,\omega)]]\}$.*

CHAPTER IV

PORTFOLIO OPTIMIZATION

4.1 Introduction

The portfolio selection problem is a classical problem in theoretical and computational finance. Historically, the seminal work of Markowitz [17, 18] which earned him the 1990 Nobel prize in economics, laid the foundation of the subject. The interest in this problem has grown over the years as the financial industry developed, along with advances in theoretical work and computing power. Various approaches and models have been proposed and elaborated to tackle this problem. Three common approaches are expected utility approach, mean/risk approach, and stochastic dominance theory.

In the expected utility approach, an investor's preference over portfolios is described quantitatively by a utility function. Utility is a notion used in economics and game theory to measure how desirable something is to the investor. It is related to, but different from, the absolute amount of financial gains. The well-known expected utility hypothesis due to Neumann and Morgenstern [22] states that for every rational decision maker there exists a utility function $u(\cdot)$ such that he prefers outcome z_1 over outcome z_2 if and only if $\mathbb{E}[u(z_1)] > \mathbb{E}[u(z_2)]$. Based on this hypothesis, the expected utility approach specifies a certain utility function $u : \mathbb{R} \to \mathbb{R}$ and formulates the following optimization problem:

$$\max_{x \in X} \mathbb{E}[u(R(x))],$$

where X denotes the set of feasible portfolios and $R(x)$ the random return of portfolio x. It is usually required that the function $u(\cdot)$ is concave and nondecreasing, thus representing preferences of a risk-averse decision maker. The choice of $u(\cdot)$ involves assumptions on the behavior of investors and therefore is somewhat subjective. So

the challenge is to select the appropriate utility function in order to yield nontrivial and meaningful solutions.

The mean/risk approach was pioneered by Markowitz's mean-variance analysis. In this approach, portfolio performance is measured in two dimensions: the mean describing the expected return rate, and the (scalar) risk representing the uncertainty of the return rate. A portfolio x is said to be *efficient* if and only if there is no other portfolio y such that $R(y)$ has higher expected value and lower risk than $R(x)$. A mean/risk model can take three forms: maximizing a weighted sum of the mean and the risk, maximizing the mean such that the risk is constrained by some specified value, or minimizing the risk such that the mean is beyond some value. Note that the first is exactly what we referred to as a mean-risk model before and can be regarded as one particular expected utility maximization model, and the second falls into the category which we called risk constrained problems. Compared with the expected utility approach, the problem structure of mean/risk models is more clear and more convenient from a computational point of view (at least for most well-known risk measures). A drawback of this approach is that sometimes the solution is not guaranteed to be meaningful with respect to the stochastic dominance relations. Recent mean/risk studies of portfolio selection problems can be found in [7, 14, 32, 33]. In all these works (except when discussing the mean/variance model), asset returns take either a discrete joint distribution or a multivariate normal distribution.

Stochastic dominance relations provide a solid theoretical basis for ranking choices, namely by pointwise comparison of some performance functions constructed from their distribution functions. The first performance function is defined as the right-continuous cumulative distribution function: $F_x^{(1)}(\eta) := \Pr\{R(x) \leq \eta\}$ where $\eta \in \mathbb{R}$. We say portfolio x dominates y under the first-order stochastic dominance (FSD) rule, $R(x) \succeq_{(1)} R(y)$, if $F_x^{(1)}(\eta) \leq F_y^{(1)}(\eta)$ for all η with at least one strict inequality. The second function is derived from the first as $F_x^{(2)}(\eta) := \int_{-\infty}^{\eta} F_x(\xi)d\xi$.

We say x dominates y under the second-order stochastic dominance (SSD) rule, $R(x) \succeq_{(2)} R(y)$, if $F_x^{(2)}(\eta) \leq F_y^{(2)}(\eta)$ for all η with at least one strict inequality. Higher-order stochastic dominance can also be studied in a similar fashion. One well-known result on stochastic dominance and expected utility is: $R(x) \succeq_{(1)} R(y)$ if and only if $\mathbb{E}[u(R(x))] \geq \mathbb{E}[u(R(y))]$ for all nondecreasing function $u(\cdot)$ and for which these expected values are finite; and $R(x) \succeq_{(2)} R(y)$ if and only if $\mathbb{E}[u(R(x))] \geq \mathbb{E}[u(R(y))]$ for all nondecreasing concave function $u(\cdot)$, for which these expected values are finite. Thus, a rational decision maker will only prefer a portfolio that is not dominated by any other portfolio in the SSD sense.

Despite its theoretical importance, stochastic dominance models are computationally very demanding. So an important study is whether certain mean/risk model is in harmony with stochastic dominance [25, 26, 27]. In a recent paper, Dentcheva and Ruszczyński [9] studied a model involving explicitly stochastic dominance constraints on the portfolio return rate. When the return rates are described by a discrete joint distribution, the model is equivalent to a linear programming problem.

In this chapter, we shall solve a portfolio optimization problem using the mean/risk approach. All our models provide stochastically nondominant solutions. Our objective includes the following: (1) test the effectiveness of the proposed SAA schemes; (2) check the effects of incorporating certain variance reduction technique into the SAA scheme for mean-MAD models; (3) compare different portfolio optimization models including three mean-risk models (mean-MAD, mean-quantile deviation, and mean-CVaR), one CVaR constrained model, and one chance constrained model; (4) compare different assumptions on the distribution of the random vector, including one discrete distribution and two continuous distributions (multivariate lognormal and multivariate normal). In Section 4.2, we give a detailed description of our portfolio problem. In Section 4.3, we explain how the data is collected and processed. In Section 4.4, we present and discuss the computation results.

4.2 Problem Description

Let $r_1, \cdots, r_K$ be random one-month dollar-per-dollar returns of assets $1, \cdots, K$. Our aim is to invest our capital in these assets and obtain some desirable characteristics of the total return on the investment. Denoting by $x_1, \cdots, x_K$ the fractions of the initial capital invested in assets $1, \cdots, K$, the total one-month loss is

$$G(x, r) = -100\ r^\top x. \tag{4.1}$$

Note that the constant 100 in (4.1) is simply for convenience of reporting the computational results and thus all returns reported are in percentage.

We consider various portfolio optimization models defined on a set P satisfying

$$\begin{aligned}
&\sum_{k=1}^{K} x_k = 1 \\
&n_l \le \sum_{k=1}^{K} y_k \le n_u \\
&y_k \in \{0, 1\}, \quad k = 1, \cdots, K \\
&v_l y_k \le x_k \le v_u y_k, \quad k = 1, \cdots, K,
\end{aligned}$$

where v_l, v_u, n_l, and n_u are constants. By introducing binary variables $y_1, \cdots, y_K$, we require that the fraction on the k-th asset is either 0 or on the interval $[v_l, v_u]$, and the total number of assets invested in should be no fewer than n_l and no more than n_u. These are resulted from considerations of lowering commission costs and diversifying assets (thus lowering risk). The associated mean-MASD model then takes the form

$$\min_{x \in P}\ \mathbb{E}[G(x, r)] + \lambda \delta_+[G(x, r)]$$

and a model minimizing the MAD for mean no larger than q has the form

$$\begin{aligned}
\min \quad & \delta[G(x, r)] \\
\text{s.t.} \quad & \mathbb{E}[G(x, r)] \le q \\
& x \in P,
\end{aligned}$$

which we call the MAD model. Similarly we can write the mean-QDEV model, the mean-CVaR model, and the variance model. We also consider the chance constrained model

$$\begin{aligned} \min \quad & \mathbb{E}[G(x,r)] \\ \text{s.t.} \quad & \Pr\{G(x,r) \le q\} \ge \alpha \\ & x \in P \end{aligned}$$

and the CVaR constrained model

$$\begin{aligned} \min \quad & \mathbb{E}[G(x,r)] \\ \text{s.t.} \quad & \text{CVaR}_\alpha[G(x,r)] \le q \\ & x \in P. \end{aligned}$$

4.3 Data Collection and Processing

In our experiments, the set of assets to invest in is 95 stocks from S&P100, excluding SBC, ATI, GS, LU and VIA-B due to insufficient data, and $v_l = 0.05$, $v_u = 0.25$, $n_l = 10$, and $n_u = 30$. All historical price data are downloaded from http://finance.yahoo.com. Stock splitting is adjusted. Several notations used subsequently are explained below.

- Lognormal, Normal and Discrete

  These refer to what distribution the vector of price ratio (i.e., return+1) takes. Here we are interested in three distributions. The first is multivariate lognormal, which is motivated by the fact that lognormal is the underlying assumption for the well-known Black-Scholes formula [8]. The second is multivariate normal, and the third is discrete. For the first two distributions, we estimate the characteristic values (mean vector and covariance matrix) from historical data. For the last one, we treat all historical data of price ratio as the scenario set and each scenario has equal probability.

- Daily and Monthly

  These refer to which kind of historical data (daily or monthly) is used. Although the uncertain parameters $r_1, \cdots, r_K$ in (4.1) are monthly returns, we might instead describe the probability distribution of the daily price ratios. In this case, we use the historical data to calculate the mean vector and covariance matrix of daily price ratio; then we generate scenarios of daily price ratios; and the product of every twenty-one independent scenarios of daily price ratios gives one scenario of monthly price ratio. On the other hand, we might assume it is the monthly price ratios whose probability distribution is specified. In this case, we directly use the historical monthly data to generate the mean vector and covariance matrix, and further do the sampling, each scenario representing one monthly price ratio.

- In-Sample and Out-of-Sample

  We download ten-year historical data from 1996 to 2005, both daily and monthly prices. We use seven-year data among them to do the in-sample optimization and obtain optimal portfolios, and the remaining three-year data to test these obtained portfolios.

- APP1, APP2 and APP3

  These refer to how we divide seven-year in-sampling and three-year out-of-sampling among the ten-year historical data. Considering that the U.S. financial market experienced both bull and bear markets in that decade, we intentionally design three time-division approaches to represent this factor. By APP1, the three-year refers to 2003, 2004 and 2005; APP2, 1999, 2001 and 2003; and APP3, 1997, 2000 and 2005.

Unless stated otherwise, the SAA parameters we use for mean-risk models are as follows: the sample size N for $\tilde{\alpha}$-estimation is 1000, the sample size N_l for lower

bounding is 1000, the sample number M_l for lower bounding is 10, and the sample size N_u for upper bounding is 100000 (see Table 2.2); and those for CVaR constrained models are: the sample size N for getting a candidate solution is 2000, the sample size N_u for testing its feasibility is 50000, the sample size N_l for lower bounding is 1000, the sample number M_l for lower bounding is 10, and the small value ξ used for decreasing the right-hand-side constant in the expected value constraint is 0.2 (see Table 3.1). And whenever we implement the SAA schemes for traditional stochastic programs and mean-MAD models, the solution $\bar{x}$ reported is the one obtained from lower bounding problems and providing the best upper bound.

4.4 Result Analysis

4.4.1 Simple Monte Carlo and Latin Hypercube Sampling Schemes

We first demonstrate the effects of variance reduction by employing the Latin Hypercube sampling (LHS) scheme [19]. The model solved is the mean-MASD model. Tables 4.1 and 4.2 give the computational results from the SAA method with simple Monte Carlo (SMC) and with LHS, respectively. In these tables, "UB 95%-Conf.Int." represents the approximate 95%-confidence interval of the true upper bound, and "LB 95%-Conf.Int." represents that of the true lower bound. We can see that with LHS employed, the standard deviation of the gap estimator $\sqrt{S_{\bar{u}}^2 + S_{\bar{l}}^2}$ is significantly reduced, e.g. a 64% reduction for $m = 1$. Consequently the confidence intervals are also greatly reduced by LHS. Henceforward, all data reported relevant to SAA are with LHS.

4.4.2 Continuous and Discrete Distributions

In this experiment, we compare the results from solving the mean-MASD model under two distribution assumptions for the uncertain parameters: one is "daily lognormal," i.e., the vector of daily price ratios of all stocks takes a multivariate lognormal distribution, and the other is "monthly discrete," i.e., the vector of monthly price ratios

Table 4.1: SMC (mean-MASD, $\lambda = 0.5$, daily lognormal, APP1)

m	$\hat{l}^m$	$\bar{u}$	$S_{\bar{u}}$	UB 95%-Conf.Int.
1	-0.6447	-0.2670	0.0301	$[-0.3260, -0.2080]$
2	-0.3854	-0.1471	0.0312	$[-0.2083, -0.0859]$
3	-0.4184	-0.2136	0.0255	$[-0.2636, -0.1636]$
4	-0.4908	-0.1574	0.0320	$[-0.2201, -0.0947]$
5	-0.8691	-0.2626	0.0296	$[-0.3206, -0.2046]$
6	-0.8794	-0.2126	0.0320	$[-0.2753, -0.1499]$
7	-0.4400	-0.3086	0.0258	$[-0.3592, -0.2580]$
8	-0.7166	-0.2457	0.0313	$[-0.3070, -0.1844]$
9	-1.1370	-0.1250	0.0432	$[-0.2097, -0.0403]$
10	-0.5707	-0.2025	0.0335	$[-0.2682, -0.1368]$
$\bar{l} = -0.6552$			$S_{\bar{l}} = 0.0776$	
LB 95%-Conf.Int.= [-0.8307, -0.4798]				

Table 4.2: LHS (mean-MASD, $\lambda = 0.5$, daily lognormal, APP1)

m	$\hat{l}^m$	$\bar{u}$	$S_{\bar{u}}$	UB 95%-Conf.Int.
1	-0.4183	-0.4403	0.0290	$[-0.4971, -0.3835]$
2	-0.4770	-0.4446	0.0293	$[-0.5020, -0.3872]$
3	-0.4550	-0.4407	0.0311	$[-0.5017, -0.3797]$
4	-0.4772	-0.4393	0.0296	$[-0.4973, -0.3813]$
5	-0.4402	-0.4325	0.0299	$[-0.4911, -0.3739]$
6	-0.4783	-0.4486	0.0316	$[-0.5105, -0.3867]$
7	-0.4496	-0.4453	0.0291	$[-0.5023, -0.3883]$
8	-0.4852	-0.4410	0.0308	$[-0.5014, -0.3806]$
9	-0.4416	-0.4512	0.0292	$[-0.5084, -0.3940]$
10	-0.4425	-0.4533	0.0297	$[-0.5115, -0.3951]$
$\bar{l} = -0.4565$			$S_{\bar{l}} = 0.0069$	
LB 95%-Conf.Int.= [-0.4722, -0.4408]				

takes a discrete distribution.

Figure 4.1 presents the efficient frontier of the mean-MASD model. The "SAA" plot denotes the efficient frontier from the "lognormal" assumption; the "H" discrete points correspond to portfolios from the "discrete" assumption, while their x-values and y-values are estimated using the sample that is used to compute the upper bound in SAA. The relative position of the "SAA" plot and the "H" points tells us that the two assumptions results in very different optimal portfolios. Furthermore, since the "discrete" distribution contains only eighty-four scenarios with equal probability

whereas the "lognormal" distribution contains, in principle, numerous scenarios with probability density information, the "SAA" portfolio dominates the "H" portfolio with respect to the lognormal distribution.

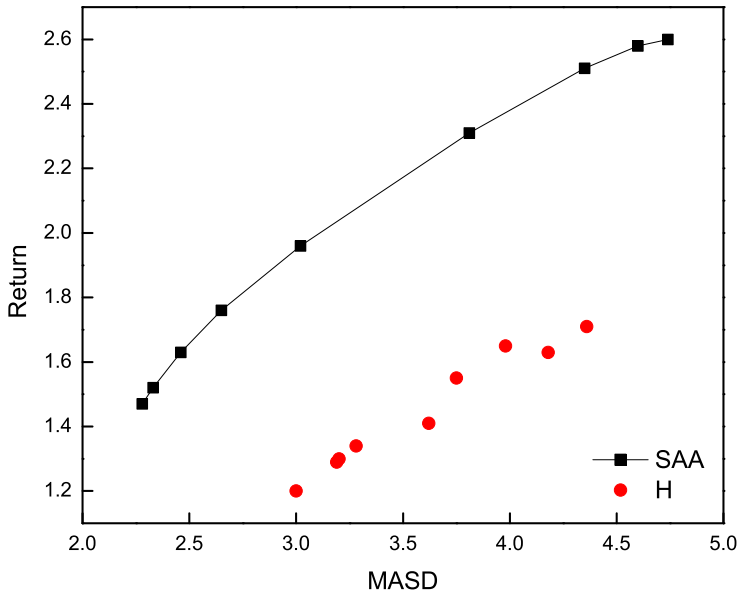


Figure 4.1: Efficient frontier (mean-MASD, APP1)

Figure 4.2 gives the composition of the optimal portfolios from the "lognormal" and "discrete" assumption. As λ increases, we can see progressive change of the portfolio composition: some stocks enter into the optimal portfolio and some leave. These charts show that more stocks are contained in a portfolio with larger λ, i.e., to reduce risk, more stocks are needed.

We also have an out-of-sample test to the optimal portfolios calculated from the mean-MASD model under nine λ-values $0.1, \cdots, 0.9$, two distribution assumptions "lognormal" and "discrete," and three time-division approaches "APP1," "APP2," and "APP3." Figure 4.3 displays the three-year profit curves those portfolios from

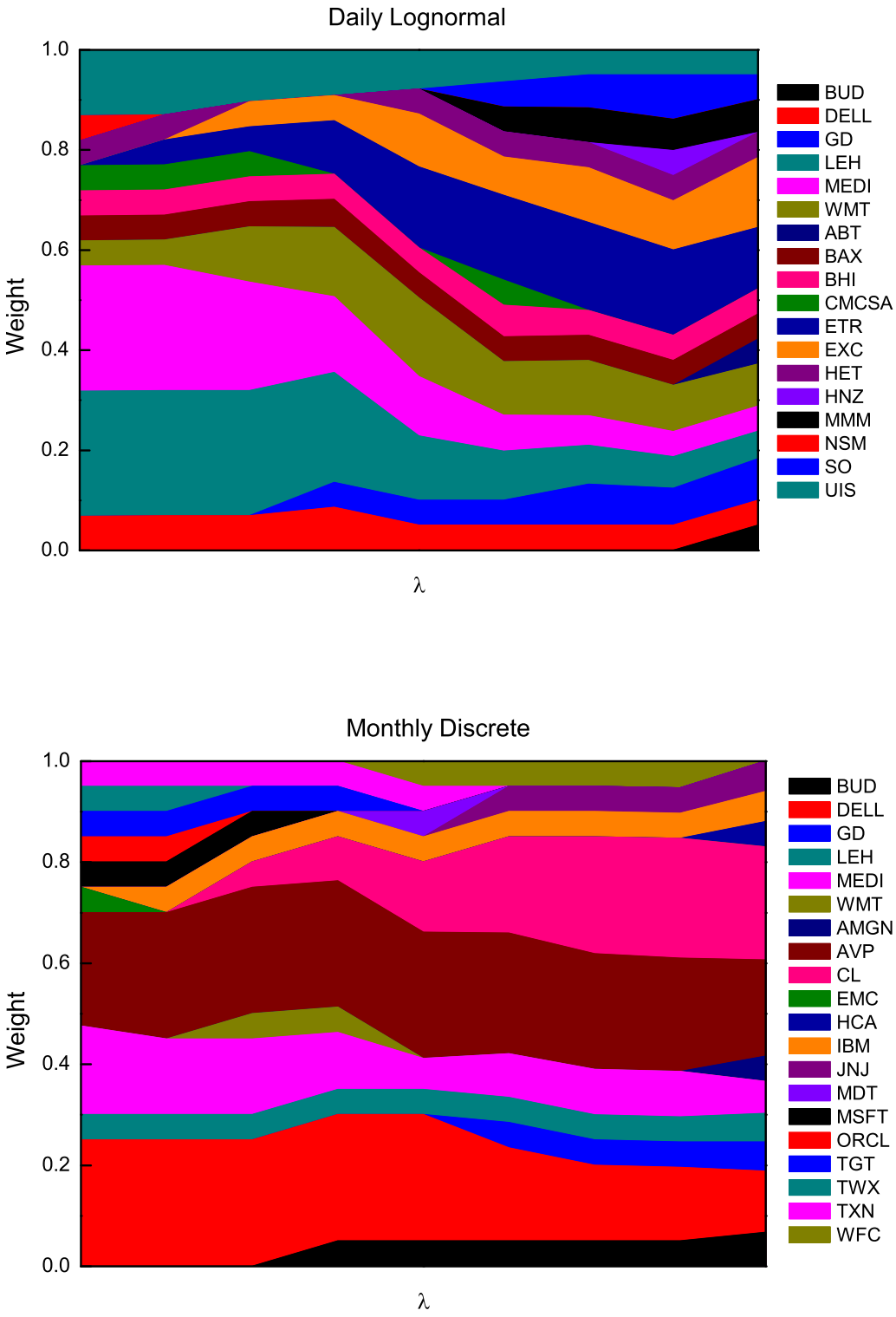


Figure 4.2: Portfolio composition (mean-MASD, APP1)

λ=0.5. Figure 4.4 displays the Sharpe ratios of the optimal portfolios from all nine λ-values. These results show that the time-division approach has a big influence on these profits. Its existence makes it difficult to draw a conclusion that one distribution assumption is better than the other. Considering the fact that the U.S. financial market experienced both bull and bear markets in these ten years, we intentionally design these three division approaches to consider this factor.

4.4.3 MAD and Variance Models

Next we compare MAD and variance. Since these two risk measures are in different units, it is inappropriate to compare the results of mean-MAD and mean-variance directly. On the other hand, it is widely known that when the return vector is multivariate normally distributed, analytically the MAD model (controlling mean and minimizing MAD) and the variance model (controlling mean and minimizing variance) should return same solutions [13]. However, numerically, we cannot expect the results from the two models to be exactly same. There are two reasons. First, the solution method of the MAD model is SAA, which only provides near-optimal solutions. Second, the solution method of the variance model (convex quadratic program) is the Barrier method, which also only returns near-optimal solutions. Our computational results do show this discrepancy.

Table 4.3 gives the computational results. The upper part corresponds to the MAD model and the lower to the variance model. In the upper part, the "Mean" row lists the mean value specified when solving the variance model, except that the first one corresponds to the problem of minimizing MAD with no constraint on mean; "Variance" is the optimal objective values; "# Stocks" is the number of stocks in the optimal portfolio; and "Time(sec.)" gives the total time consumed in seconds. In the lower part, "Mean," "# Stocks," and "Time(sec.)" have the similar meaning as above; "UB" and "LB" are the upper and lower bounds returned by the SAA

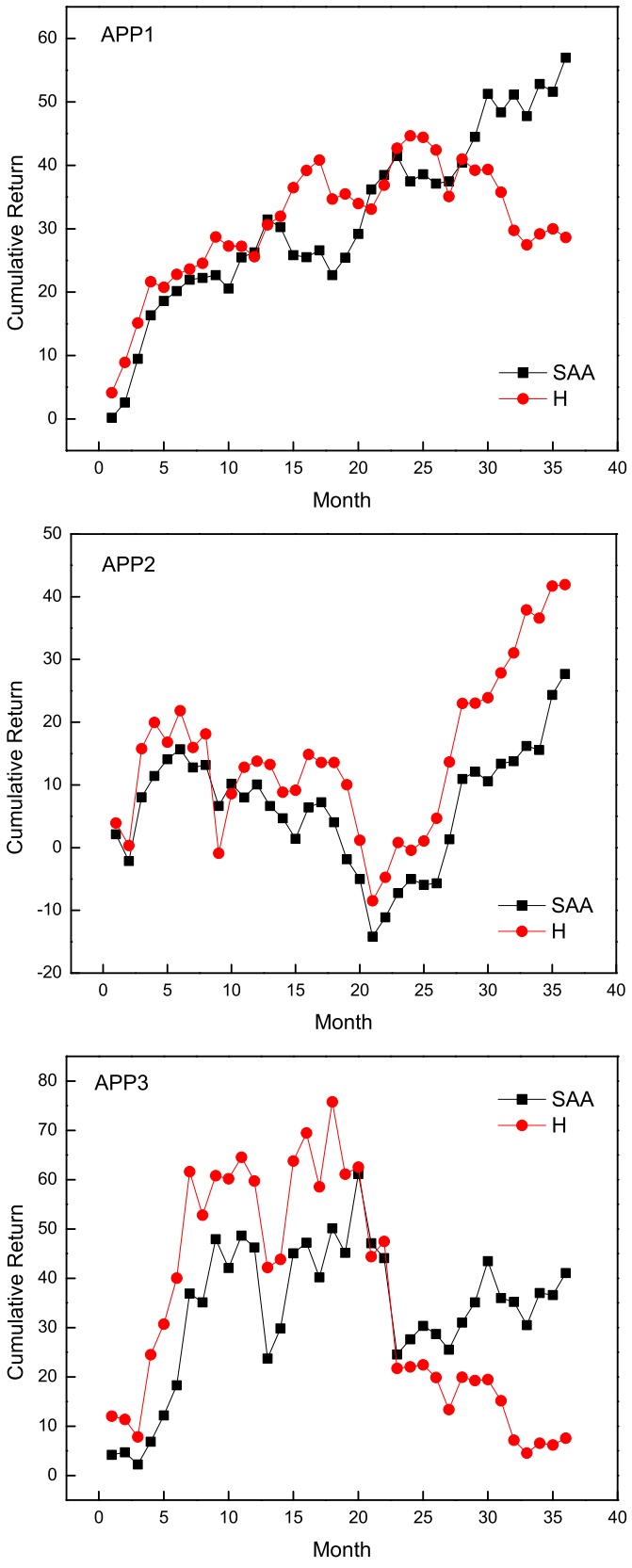


Figure 4.3: Three-year profits (Portfolio, mean-MASD, $\lambda = 0.5$)

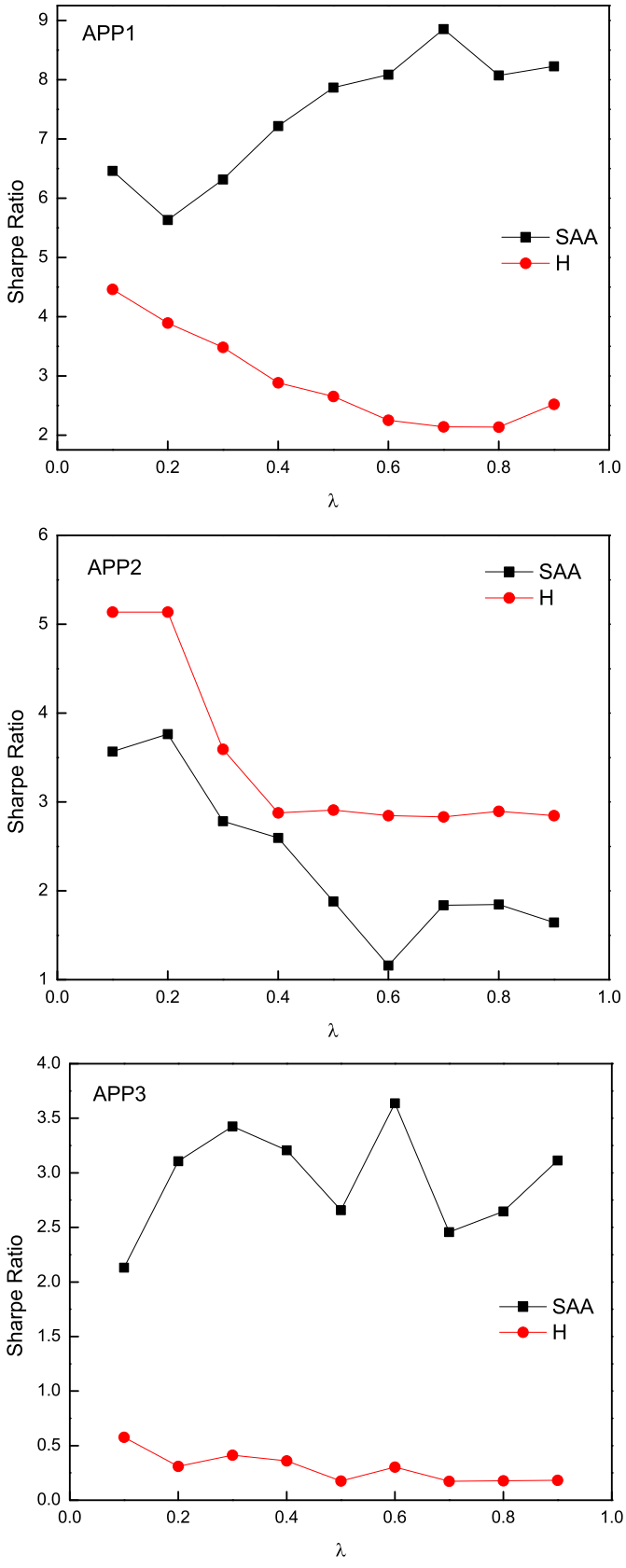


Figure 4.4: Sharpe ratios (mean-MASD)

method, and "UBSD" and "LBSD" are their standard deviations; "$\frac{\text{UB-LB}}{\text{UB}}(\%)$" is the percentage of the optimality gap; and finally "Variance" is the variance value of the optimal solution obtained in the MAD model. In particular, we draw the plots of "Time(sec.)" versus "mean" in Figure 4.5. We also provide the composition of the optimal portfolios in Figure 4.6, which demonstrates the similarity of solutions from the two models.

We know that magnitude of the expected return controls size of the feasible region of these two models. As the expected return increases, the feasible region shrink, along with several effects worth discussing. The first is on computation time. From Figure 4.5, we see that once the expected return goes beyond some value (for example, 0.015), the computation time for either model decreases significantly. The second is on portfolio composition. The number of stocks "# Stocks" contained in the optimal portfolio tends to decrease. The third effect is on performance of SAA for solving the MAD model. The relative optimality gap "$\frac{\text{UB-LB}}{\text{UB}}(\%)$" decreases, i.e., SAA returns an optimal solution with improving quality.

4.4.4 Mean-MASD, Mean-QDEV and Mean-CVaR Modes

In this experiment, we compare results from various mean-risk models: mean-MASD, mean-QDEV, and mean-CVaR. We solve the first model by the SAA method proposed here, and the other two by the SAA method for traditional stochastic programs.

Table 4.4 gives the computation time of the three models at λ=0.5. In this table, "LB-total" represents the time consumed for lower bounding; "UB-sampling" represents the time taken for generating the sample for upper bounding, and since Latin Hypercube sampling is adopted, it takes pretty long time (around 370 seconds) to generating a sample of size 10^5; and "UB-total" is the time for upper bounding, which includes the "UB-sampling." We can see that the upper bounding of mean-QDEV and mean-CVaR takes much more time than mean-MASD, because they require a sorting

Table 4.3: Performance of MAD and variance models (monthly normal, APP1)

The Variance Model											
Mean	-0.992	-1.0	-1.2	-1.4	-1.6	-1.8	-2.0	-2.2	-2.4	-2.6	-2.8
Variance	5.887	5.889	5.996	6.456	7.235	8.366	10.224	12.856	16.651	21.694	29.186
# Stocks	15	15	16	16	16	14	12	12	10	10	10
Time(sec.)	0.71	0.69	0.59	0.99	0.71	0.41	0.21	0.15	0.14	0.12	0.12
The MAD Model											
Mean	-0.955	-1.0	-1.2	-1.4	-1.6	-1.8	-2.0	-2.2	-2.4	-2.6	-2.8
UB	1.981	1.970	1.984	2.054	2.180	2.346	2.581	2.897	3.288	3.748	4.355
UBSD($\times 10^2$)	0.473	0.471	0.474	0.491	0.520	0.561	0.617	0.693	0.785	0.895	1.039
LB	1.913	1.916	1.938	2.006	2.117	2.289	2.540	2.864	3.267	3.741	4.328
LBSD($\times 10^2$)	0.683	0.680	0.580	0.666	0.598	0.871	1.372	2.357	3.209	3.933	4.933
$\frac{\text{UB-LB}}{\text{UB}}(\%)$	3.47	2.75	2.32	2.37	2.89	2.40	1.57	1.16	0.62	0.19	0.63
Variance	6.167	6.099	6.170	6.647	7.456	8.662	10.470	13.181	16.974	22.052	29.742
# Stocks	16	15	15	16	15	14	13	12	11	10	10
Time(sec.)	700	679	770	754	434	233	155	99	90	60	54

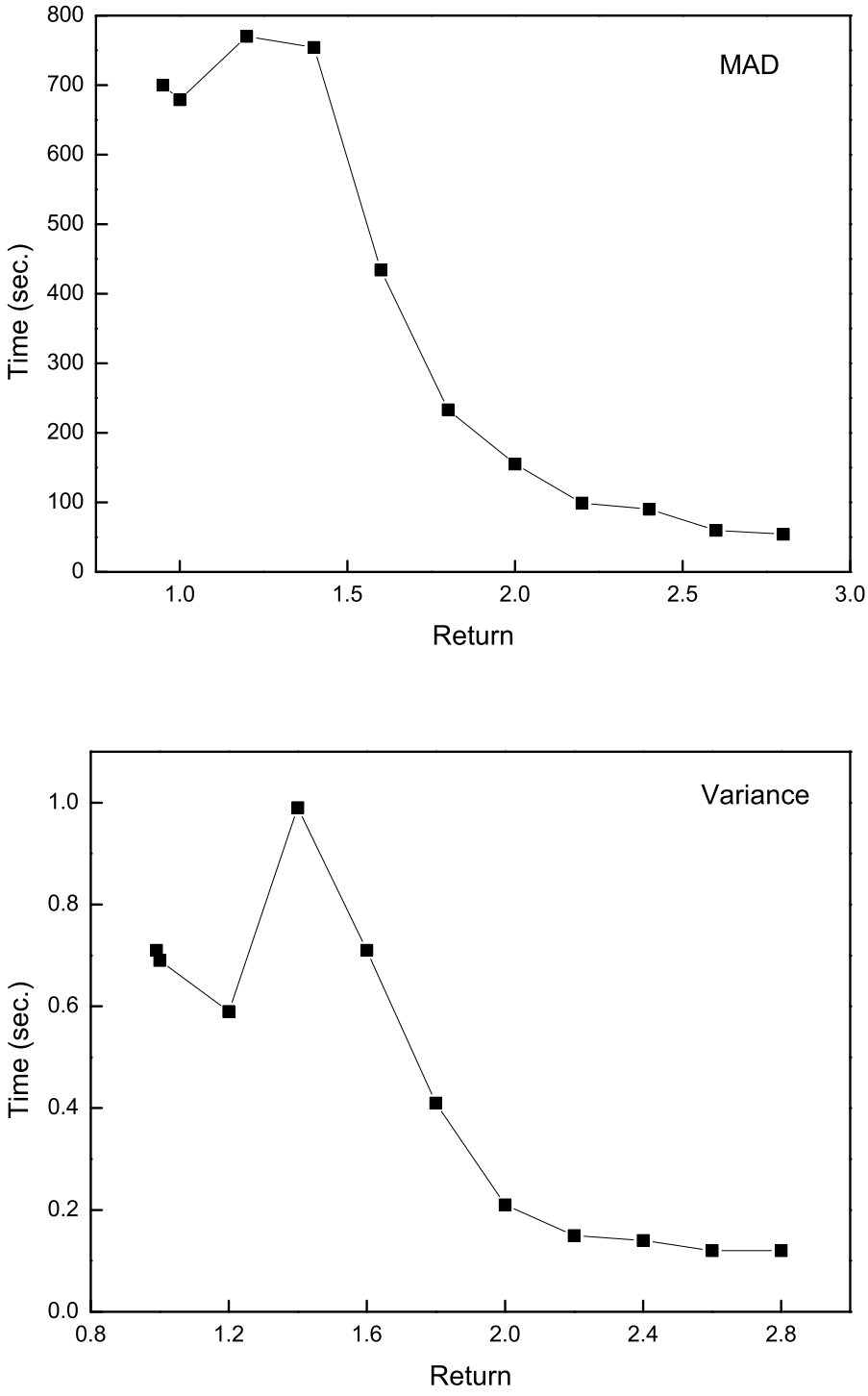


Figure 4.5: Computation time (monthly normal, APP1)

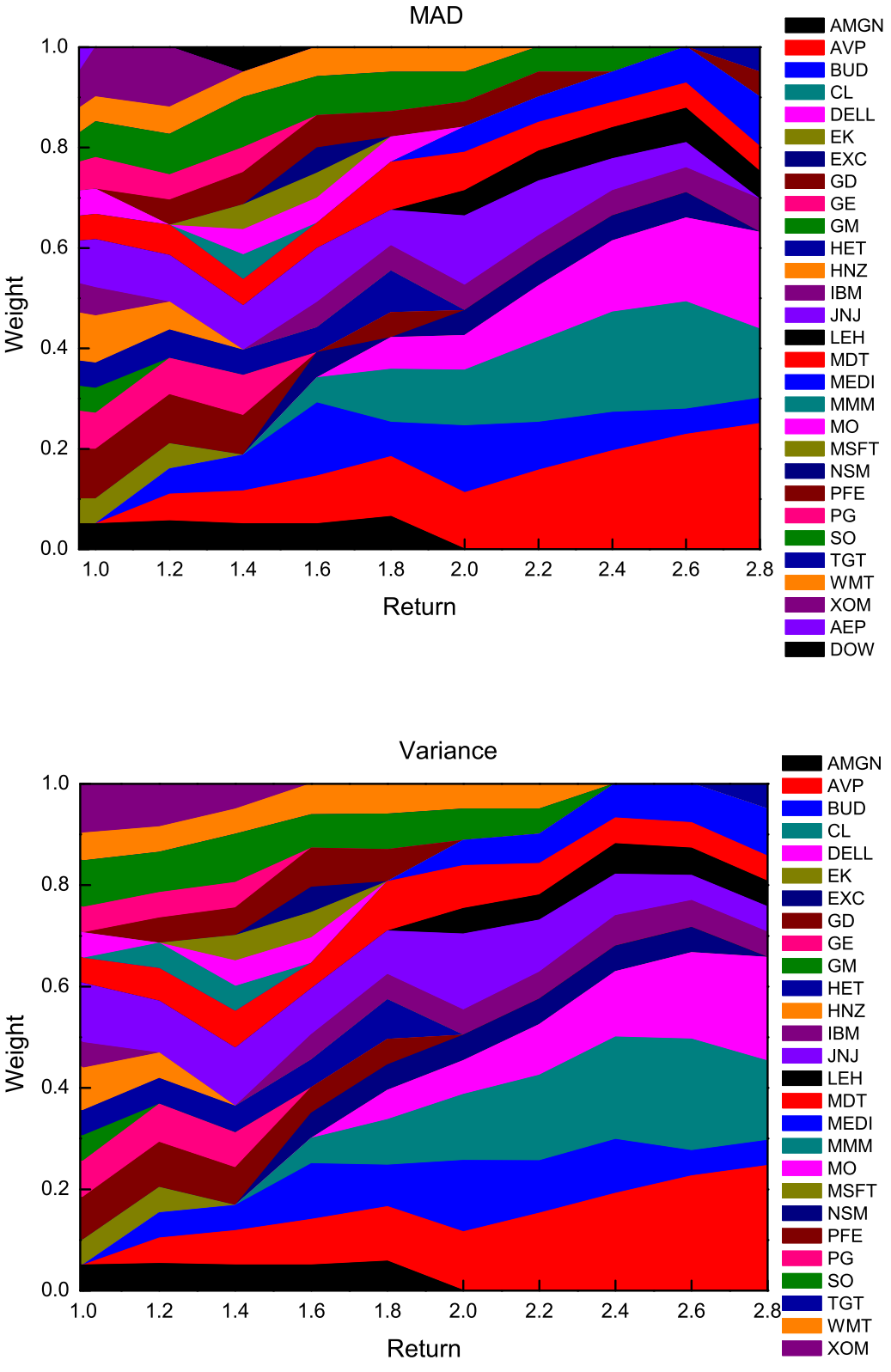


Figure 4.6: Portfolio composition (monthly normal, APP1)

procedure to get α-quantile and α-VaR, which is time-consuming, and mean-MASD only needs to compute the average.

Figure 4.7 gives the three-year profit curves from out-of-sample tests on the optimal portfolios obtained from the these models. We can see that these portfolios are pretty similar.

Table 4.4: Computation time of mean-risk models (λ=0.5, daily lognormal, APP1)

Time(sec.)	α-estimating	LB-total	UB-sampling	UB-total	total
mean-QDEV, α=0.7	N/A	79	361	1607	1686
mean-CVaR, α=0.7	N/A	289	389	1666	1955
mean-MAD	9	139	370	372	519

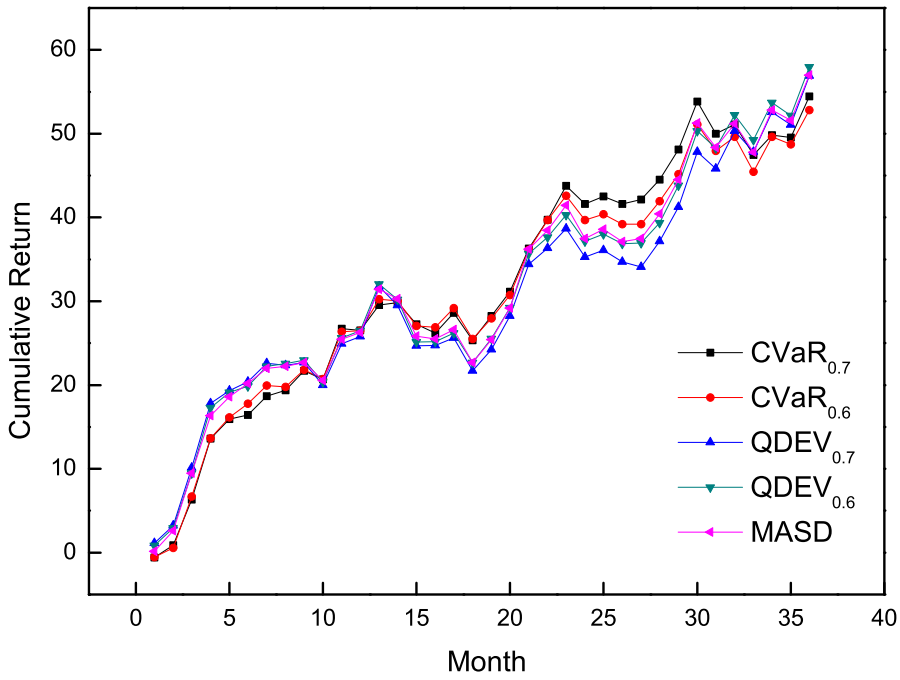


Figure 4.7: Three-year profits (daily lognormal, APP1)

4.4.5 Chance Constrained and CVaR Constrained Models

Tables 4.5 and 4.6 give the computational results of three models with respect to changing α and q, respectively. The three models are chance constrained model under multivariate normal assumption (solved by SOCP solver), CVaR constrained model under multivariate normal assumption (SAA) and CVaR constrained model under multivariate lognormal assumption (SAA). In these tables, "Opt.Obj." is the optimal objective value of the chance constrained problem; "Pr{UB}" is the estimated probability that "UB" is a statistical upper bound of the CVaR constrain problem, or equivalently, the estimated probability that the optimal solution $\tilde{x}$ corresponding to "UB" is feasible to the CVaR constrained problem; "95%-C.I.(LB)-l" and "95%-C.I.(LB)-r" are the left and right end of the 95% confidence interval of the lower bound "LB;" "Pr$\{G(\tilde{x}, r) \leq q\}(\%)$" is the true probability that $\tilde{x}$ is feasible to the chance constrained problem (the probability can be calculated due to the fact that $G(\tilde{x}, r)$ is a normal variable); "# Iterations" is the iterations taken for getting a statistical feasible solution (Step 1 in SAA for EVC), and finally as the name suggests, "UB Time," "LB Time," and "Time" are the times in seconds for upper bounding, lower bounding and the whole SAA implementation, respectively. Furthermore, we plot the objective values of the three models in Figure 4.8.

We make the following observations. The first is regarding the two assumptions used in the CVaR constrained problem. We can see that "UB-Normal" is always larger than "UB-Lognormal."

The second is regarding the SAA performance. In our SAA implementation, we set 97.7% as the smallest acceptable probability that the obtained optimal portfolio is feasible to the CVaR constrained problem. It turns out that the actual probability "Pr{UB}" returned by our codes is much higher, for example, over 99.5% in eight out of ten runs of the CVaR constrained problem with normal distribution. Moreover, we find that as α increases, number of iterations "# Iterations" for upper bounding

increases.

The third is regarding the computation time. We notice that as q increases or α decreases, it takes less time for SAA to do the upper and lower bounding. Here we give one explanation. The method we used for solving upper and lower bounding SAA problems is the cutting plane method, whose computation time is closely related on number of cuts being added. The less cuts, the less computation time. For an upper bounding problem, as q increases or α decreases, its feasible region expands and thus less cuts are added before we reach the optimality. For a lower bounding problem, as q increases or α decreases, the obtained Lagrangian multiplier $\tilde{\pi}$ decreases, which results in more smoothness of the objective function of this lower bounding problem, and hence also less cuts are needed.

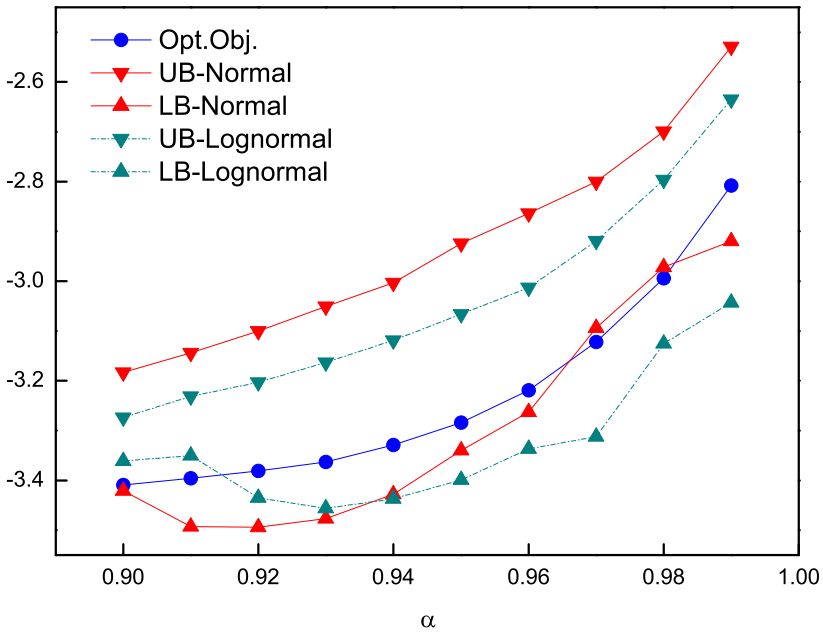
The last is regarding the approximation of a chance constraint by a CVaR one. In general, for the problem we study here, we are pretty satisfied with the approximation. Relative gaps $\frac{\text{Opt.Obj}-\text{UB}}{\text{Opt.Obj.}}$ and $\frac{\Pr\{G(\tilde{x},r)\leq q\}-\alpha}{\alpha}$ are two indicators of the approximation effect. Our intuition is larger gaps indicate more conservativeness in the approximation. However, these relative gaps are also influenced by the SAA performance. So we shall only make some observation and draw no conclusions. As α (q) increases, $\frac{\Pr\{G(\tilde{x},r)\leq q\}-\alpha}{\alpha}$ $\left(\frac{\text{Opt.Obj}-\text{UB}}{\text{Opt.Obj.}}\right)$ keeps decreasing. However, there is no obvious correlation between q (α) and $\frac{\Pr\{G(\tilde{x},r)\leq q\}-\alpha}{\alpha}$ $\left(\frac{\text{Opt.Obj}-\text{UB}}{\text{Opt.Obj.}}\right)$.

Table 4.5: Effects of α ($q = 10$, monthly, APP1)

α(%)	90	91	92	93	94	95	96	97	98	99
The Chance Constrained Model (Normal)										
Opt.Obj.	-3.409	-3.396	-3.381	-3.363	-3.329	-3.284	-3.219	-3.122	-2.994	-2.808
Time(sec.)	0.29	0.30	0.32	0.30	14.67	3.30	6.58	21.25	12.64	9.37
The CVaR Constrained Model (Normal)										
UB	-3.183	-3.144	-3.100	-3.051	-3.003	-2.924	-2.864	-2.800	-2.699	-2.529
Pr{UB}(%)	100.00	100.00	100.00	100.00	97.83	100.00	99.97	99.92	99.67	98.97
LB	-3.421	-3.493	-3.494	-3.477	-3.428	-3.340	-3.263	-3.094	-2.972	-2.920
LBSD($\times 10^2$)	0.750	0.846	0.853	0.866	0.897	0.914	0.886	0.790	1.047	1.324
95%-C.I.(LB)-l	-3.438	-3.512	-3.514	-3.496	-3.449	-3.361	-3.283	-3.111	-2.996	-2.950
95%-C.I.(LB)-r	-3.404	-3.474	-3.475	-3.457	-3.408	-3.319	-3.243	-3.076	-2.948	-2.890
$\Pr\{G(\tilde{x}, r) \leq q\}$(%)	96.41	96.72	97.15	97.50	97.77	98.28	98.60	98.97	99.32	99.68
$\frac{\Pr\{G(\tilde{x},r)\leq q\}-\alpha}{\alpha}$(%)	7.12	6.29	5.60	4.84	4.01	3.45	2.71	2.03	1.35	0.68
$\frac{\text{Opt.Obj}-\text{UB}}{\text{Opt.Obj.}}$(%)	6.61	7.40	8.31	9.29	9.79	10.97	11.01	10.32	9.87	9.94
UB Time(sec.)	122.46	125.92	132.23	127.85	142.92	185.52	182.50	183.18	228.67	309.15
LB Time(sec.)	30.83	34.81	54.10	54.92	60.63	55.62	61.08	55.97	65.00	63.88
Total Time(sec.)	153.29	160.73	186.33	182.77	203.55	241.14	243.58	250.19	293.67	373.03
# Iterations	2	2	2	2	2	3	3	3	4	6
$\tilde{\lambda}$	0.184	0.211	0.223	0.231	0.234	0.228	0.228	0.206	0.200	0.217
The CVaR Constrained Model (Lognormal)										
UB	-3.274	-3.231	-3.203	-3.163	-3.119	-3.066	-3.013	-2.919	-2.796	-2.635
Pr{UB}(%)	99.90	100.00	100.00	100.00	100.00	99.96	99.81	100.00	100.00	100.00
LB	-3.361	-3.350	-3.435	-3.456	-3.437	-3.399	-3.337	-3.312	-3.125	-3.043
LBSD($\times 10^2$)	0.844	0.765	0.699	0.799	0.841	0.857	0.865	1.059	1.169	1.536
95%-C.I.(LB)-l	-3.380	-3.367	-3.451	-3.474	-3.456	-3.419	-3.356	-3.336	-3.152	-3.077
95%-C.I.(LB)-r	-3.342	-3.333	-3.419	-3.438	-3.418	-3.380	-3.317	-3.288	-3.099	-3.008
UB Time(sec.)	80.77	115.55	121.89	127.34	131.82	132.30	126.80	180.33	269.67	314.20
LB Time(sec.)	20.68	21.40	29.40	29.91	42.55	43.81	50.08	56.64	54.87	57.73
Time(sec.)	101.45	136.95	151.29	157.25	174.37	176.11	176.88	236.97	324.54	371.93
# Iterations	1	2	2	2	2	2	2	3	5	6
$\tilde{\lambda}$	0.137	0.147	0.186	0.204	0.213	0.220	0.221	0.236	0.213	0.223

Table 4.6: Effects of q ($\alpha = 95\%$, monthly, APP1)

q	5	6	7	8	9	10	11	12	13	14
The Chance Constrained Model (Normal)										
Opt.Obj.	-2.553	-2.766	-2.932	-3.077	-3.199	-3.284	-3.340	-3.375	-3.395	-3.412
Time(sec.)	3.66	10.01	15.16	8.58	2.95	3.32	4.60	0.31	0.29	0.27
The CVaR Constrained Model (Normal)										
UB	-2.015	-2.268	-2.484	-2.689	-2.822	-2.924	-3.060	-3.165	-3.245	-3.300
Pr{UB}(%)	99.91	100.00	100.00	99.92	100.00	100.00	99.95	99.94	99.09	100.00
LB	-2.352	-2.646	-2.851	-3.000	-3.209	-3.340	-3.427	-3.518	-3.312	-3.335
LBSD($\times 10^2$)	1.593	1.357	1.113	1.012	1.000	0.914	0.764	0.704	1.198	1.069
95%-C.I.(LB)-l	-2.388	-2.676	-2.876	-3.022	-3.232	-3.361	-3.444	-3.534	-3.339	-3.359
95%-C.I.(LB)-r	-2.316	-2.615	-2.826	-2.977	-3.187	-3.319	-3.410	-3.502	-3.285	-3.311
$\Pr\{G(\tilde{x}, r) \leq q\}(\%)$	98.22	98.39	98.31	98.22	98.26	98.28	98.24	98.23	98.17	98.26
$\frac{\Pr\{G(\tilde{x}, r) \leq q\} - \alpha}{\alpha}(\%)$	3.39	3.57	3.49	3.38	3.44	3.45	3.41	3.40	3.34	3.43
$\frac{\text{Opt.Obj} - \text{UB}}{\text{Opt.Obj.}}(\%)$	21.10	17.98	15.26	12.61	11.78	10.97	8.39	6.24	4.42	3.27
UB Time(sec.)	246.89	297.68	246.04	145.16	143.56	187.94	127.89	122.70	117.10	115.29
LB Time(sec.)	123.78	98.55	82.38	70.45	64.56	64.50	52.01	36.44	20.72	14.52
Time(sec.)	370.67	396.23	328.42	215.61	208.12	252.44	179.90	159.14	137.82	129.81
# Iterations	2	3	3	2	2	3	2	2	2	2
$\tilde{\lambda}$	0.549	0.410	0.329	0.272	0.256	0.228	0.201	0.181	0.102	0.068
The CVaR Constrained Model (Lognormal)										
UB	-2.110	-2.357	-2.623	-2.791	-2.934	-3.066	-3.181	-3.268	-3.321	-3.358
Pr{UB}(%)	99.99	100.00	99.26	100.00	100.00	99.96	100.00	99.89	99.98	100.00
LB	-2.469	-2.758	-2.949	-3.163	-3.355	-3.399	-3.465	-3.328	-3.363	-3.385
LBSD($\times 10^2$)	1.681	1.475	1.278	1.254	1.164	0.857	0.680	0.871	0.808	0.543
95%-C.I.(LB)-l	-2.507	-2.791	-2.978	-3.192	-3.381	-3.419	-3.480	-3.348	-3.381	-3.397
95%-C.I.(LB)-r	-2.431	-2.724	-2.920	-3.135	-3.329	-3.380	-3.449	-3.308	-3.344	-3.373
UB Time(sec.)	219.29	286.72	150.96	136.01	131.97	132.67	121.27	111.97	114.87	103.75
LB Time(sec.)	111.73	88.00	61.16	58.04	56.96	43.84	28.28	17.65	13.34	11.18
Time(secs)	331.02	374.72	212.12	194.05	188.93	176.51	149.55	129.62	128.21	114.93
# Iterations	2	3	2	2	2	2	2	2	2	2
$\tilde{\lambda}$	0.582	0.428	0.335	0.297	0.270	0.220	0.186	0.106	0.074	0.046



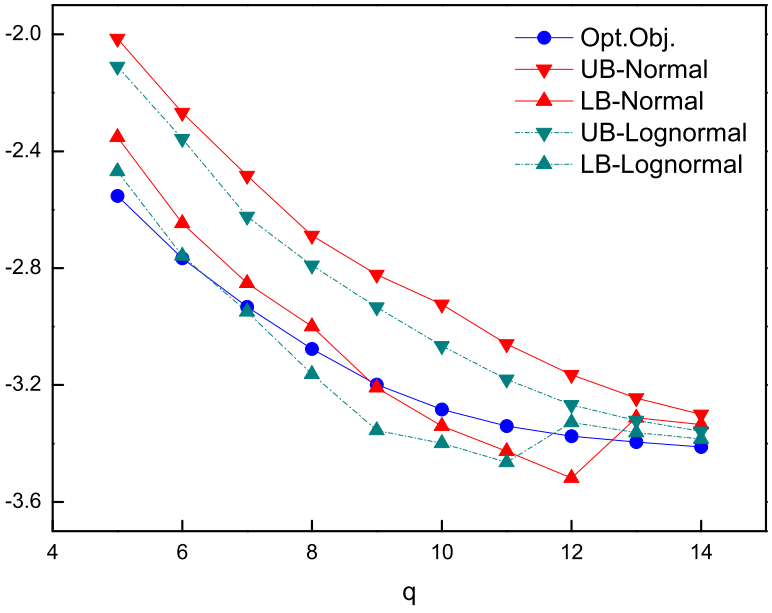


Figure 4.8: Effects of α and q (monthly, APP1)

CHAPTER V

SUPPLY CHAIN NETWORK DESIGN UNDER UNCERTAINTY

5.1 Introduction

A critical component of the planning activities of any manufacturing firm is the efficient design and operation of its supply chain. Generally a supply chain consists of a network of suppliers, manufacturing plants, warehouses, and distribution centers organized to acquire raw materials, convert them to final products, and deliver the products to customers. The design decision of a supply chain network is made on two levels: the strategic level and the tactical level. The strategic level design involves deciding the configuration of the network, i.e., the number, location, capacity, and technology of the facilities. The tactical level design involves deciding the operation of the network, i.e., controlling material flows for purchasing, processing, and distribution of products.

The impact of the design decisions spans a long horizon, during which many parameters such as costs, demands, and distances may fluctuate. For this reason, the solutions of design models without considering such fluctuation are usually far from optimal in the long run. The two-stage nature of this problem (choose locations before knowing the future demands, and react once uncertainties are resolved) makes it a classical example of decision making under uncertainty. A larger number of approaches for optimization under uncertainty have been applied to the supply chain design problem. Snyder [43] gave a thorough literature review on this subject, where the author listed a variety of stochastic and robust models for this problem. Among

the stochastic models, most extensively studied are traditional stochastic programming models. By comparison, there are only a few risk-averse stochastic programming models, including mean-variance, chance-constrained, and max-probability models.

Two earlier papers are directly relevant to our work in this chapter. Tsiakis et al. [45] considered a multi-product, multi-echelon supply chain under demand uncertainty. The demand vector has three scenarios. The design goal is to decide middle-echelon facility locations and capacities, transportation links, and flows to minimize expected cost. Transportation costs are piecewise linear concave. They formulated a large-scale mixed-integer program and solved it using CPLEX. We shall consider a supply chain network with similar (but not the same) physical structures, and our mathematical formulation of the design problem bears some resemblance to that of [45]. Santoso et al.[36] investigated a global supply chain network design problem with continuously distributed random costs, demands, and capacities. The goal is to decide the location of facilities and the type of machines to build at each facility so as to minimize the total expected cost. In case that the constructed capacity is insufficient to supply the realized demand, the model includes a shortfall penalty. The authors formulated this problem as a traditional two-stage stochastic program with binary variables in the first stage and continuous variables in the second stage and solved it using sample average approximation combined with accelerated Benders decomposition. Our solution strategy will be conceptually parallel to that of [36].

In this chapter we study a multi-echelon supply chain network design problem under demand and cost uncertainty which is continuously distributed. We investigate three optimization models: a deterministic model, a traditional stochastic model, and a mean-MASD stochastic model. To solve the stochastic models, which are two-stage stochastic programs with mixed-integer first-stage variables and continuous second-stage variables, we use appropriate sample average approximation methods to transform the original stochastic problems to a series of deterministic problems,

and then solve these deterministic problems. In Section 5.2, we describe the multi-echelon supply chain network design problem and introduce the deterministic model. In Section 5.3, we introduce the stochastic models. Finally in Section 5.4, we present the computational results.

5.2 Problem Description and Deterministic Model

This work considers the design of multi-product, multi-echelon production and distribution networks. As shown in Figure 5.1, the network consists of a number of existing multi-product manufacturing sites at fixed locations, a number of warehouses and distribution centers of unknown locations (to be selected from a set of candidate locations), and a number of customer zones at fixed locations. In general, each product can be produced at several plants located at different locations. Warehouses can be supplied from more than one manufacturing site and can supply more than one distribution center. Each distribution center can be supplied by more than one warehouse. Each customer zone places demands for one or more products, which is fulfilled by distribution centers.

The supply chain configuration decisions consist of deciding the number, location, and capacity of warehouses and distribution centers to be established. The operational decisions include the flow and production rate of materials. The objective is to minimize the total cost of the network under the premise that the customer demands are satisfied. The total cost includes a fixed infrastructure part, incurred by the establishment of warehouses and distribution centers, and an operational part, associated with production, handling of products at warehouses and distribution centers, and transportation. Before we build the deterministic model, we first give the definition of the variables and describe various conditions to be satisfied by any feasible solution. We follow partly [45] for the basic structure of the network as well as the mathematical formulation of the physical problem. In their paper, one deterministic

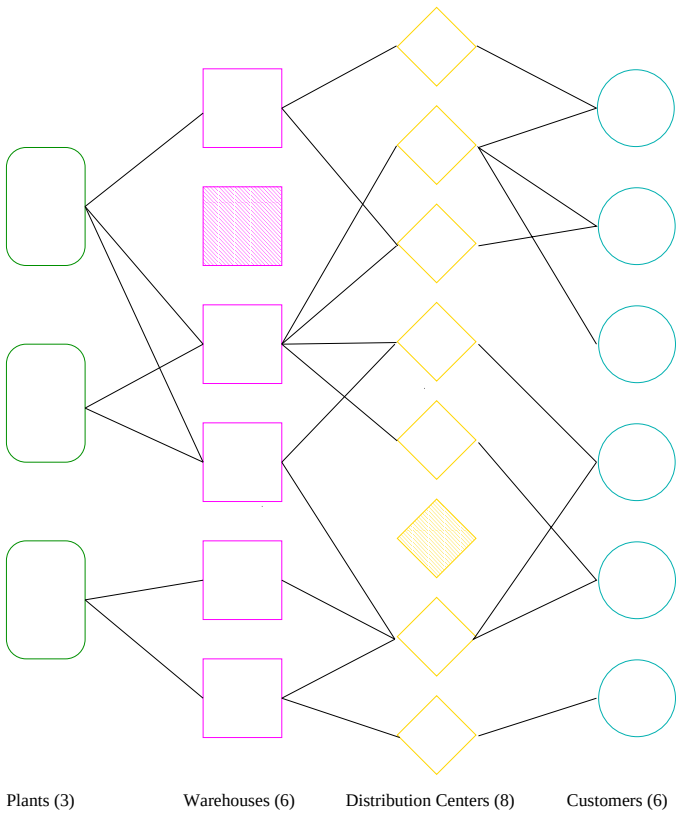


Figure 5.1: Diagram of a supply chain network

model and one stochastic model with three demand scenarios are considered.

5.2.1 Variables

Table 5.1 gives the definition of all variables used in our models, where X_m, X_k, W_m, and D_k are first-stage (or configurational) variables, and P_{ij}, Y_{ijm}, Y_{imk}, and Y_{ikl} second-stage (or operational) variables. The optimal value of operational variables will depend strongly on production cost, demands, and network configuration. In the deterministic model, we assume all parameters are deterministic. Later on, we shall consider the cases where product demands and production costs are uncertain with known distribution. Note that the parameters uncertain in stochastic models take their mean values in the deterministic model.

Table 5.1: Variable definition

i	subscript for product
j	subscript for plant
m	subscript for warehouse
k	subscript for distribution center
l	subscript for customer zone
e	subscript for resource
X_m	binary variable, 1 (0) if warehouse m is (not) built
X_k	binary variable, 1 (0) if distribution center k is (not) built
W_m	the capacity of warehouse m
D_k	the capacity of distribution center k
P_{ij}	the production rate of product i at plant j
Y_{ijm}	the flow of product i from plant j to warehouse m
Y_{imk}	the flow of product i from warehouse m to distribution center k
Y_{ikl}	the flow of product i from distribution center k to customer zone l

5.2.2 Constraints

Now we describe various conditions to be satisfied by our model. The meaning of some constraints is straightforward.

- Logical constraints on flow capacities

$$Y_{ijm} \leq Q_{ijm}^{\max} X_m, \quad \forall i, j, m \tag{5.1}$$

$$Y_{imk} \leq Q_{imk}^{\max} X_m, \quad \forall i, m, k \tag{5.2}$$

$$Y_{imk} \leq Q_{imk}^{\max} X_k, \quad \forall i, m, k \tag{5.3}$$

$$Y_{ikl} \leq Q_{ikl}^{\max} X_k, \quad \forall i, k, l \tag{5.4}$$

  where $Q_{ijm}^{\max}$ stands for the flow capacity of product i from plant j to warehouse m, similarly for $Q_{imk}^{\max}$ and $Q_{ikl}^{\max}$.

- Material balance constraints

$$P_{ij} = \sum_m Y_{ijm} \quad \forall i, j \tag{5.5}$$

$$\sum_j Y_{ijm} = \sum_k Y_{imk} \quad \forall i, m \tag{5.6}$$

$$\sum_m Y_{imk} = \sum_l Y_{ikl} \quad \forall i, k \tag{5.7}$$

$$\sum_k Y_{ikl} \geq D_{il} \quad \forall i, l \tag{5.8}$$

In above, D_{il} stands for the demand of customer zone l on product i.

- Production resource constraints

$$\sum_i \rho_{ije} P_{ij} \leq R_{je} \quad \forall j, e \tag{5.9}$$

where ρ_{ije} expresses the amount of resource e consumed by plant j to produce a unit of product i, and R_{je} the total availability of resource e at plant j.

- Capacity constraints of warehouses and distribution centers

$$W_m \leq W_m^{\max} X_m \quad \forall m \tag{5.10}$$

$$D_k \leq D_k^{\max} X_k \quad \forall k \tag{5.11}$$

$$W_m \geq \sum_{ik} \alpha_{im} Y_{imk} \quad \forall m \tag{5.12}$$

$$D_k \geq \sum_{il} \beta_{il} Y_{ikl} \quad \forall k \tag{5.13}$$

where $W_m^{\max}$ and $D_k^{\max}$ represent the upper bounds on the capacity of warehouse m and distribution center k. The latter two come from the general assumption that capacity of any warehouse or distribution center is related linearly to the flows of materials that it handles. α_{im} and β_{il} are given coefficients (equal to 1 in our experiments).

- Variable domain

$$X_m, X_k \in \{0,1\} \quad \forall m,k \tag{5.14}$$

$$W_m, D_k \geq 0 \quad \forall m,k \tag{5.15}$$

$$P_{ij}, Y_{ijm}, Y_{imk}, Y_{ikl} \geq 0 \quad \forall i,j,m,k,l \tag{5.16}$$

5.2.3 Objective Function

The overall objective is

$$\begin{aligned}\min \quad & \sum_m C_m^{\mathrm{B}} X_m + \sum_k C_k^{\mathrm{B}} X_k + \sum_m C_m^{\mathrm{U}} W_m + \sum_k C_k^{\mathrm{U}} D_k \\ & + \sum_{ij} C_{ij}^{\mathrm{P}} P_{ij} + \sum_{im} C_{im}^{\mathrm{H}} \sum_j Y_{ijm} + \sum_{ik} C_{ik}^{\mathrm{H}} \sum_m Y_{imk} \\ & + \sum_{ijm} C_{ijm}^{\mathrm{T}} Y_{ijm} + \sum_{imk} C_{imk}^{\mathrm{T}} Y_{ijm} + \sum_{ikl} C_{ikl}^{\mathrm{T}} Y_{ikl}\end{aligned}$$

where C_m^{B} (C_k^{B}) represents the fixed building cost of warehouse m (distribution center k); C_m^{U} (C_k^{U}) is the building cost corresponding to one unit of capacity of warehouse m (distribution center k); C_{ij}^{P} the unit production cost of product i at plant j; C_{im}^{H} (C_{ik}^{H}) denotes the unit handling cost of product i at warehouse m (distribution center k); C_{ijm}^{T} (C_{imk}^{T}, C_{ikl}^{T}) is the unit transportation cost of product i from plant j to warehouse m (warehouse m to distribution center k, distribution center k to customer zone l). Note that the first four terms in the objective represent configuration cost and the left terms operational cost.

Overall, this formulation gives a deterministic, mixed-integer program, which can be solved efficiently by any existing mixed-integer programming algorithm.

5.3 Supply Chain Network Design under Uncertainty

The formulation presented in the previous section assumes that all parameters, e.g. production cost, handling cost, transportation cost as well as product demand, are deterministic. However in a real supply chain, often these are uncertain. In our study, we assume there exist uncertainties in production costs and customer demands. To

design a supply chain network capable of handling these uncertainties, we adopt stochastic programming approaches and build two stochastic models: traditional SP model and risk-averse mean-MASD model.

Let us write down the two formulations. The traditional SP model takes the following form:

$$\begin{aligned} \min \quad & f(x) + \mathbb{E}[G(x,\omega)] \\ \text{s.t.} \quad & x \in X \end{aligned} \tag{5.17}$$

where

$$\begin{aligned} G(x,\omega) = \quad & \min \quad h(y) \\ & \text{s.t.} \quad y \in Y(x,\omega). \end{aligned}$$

$$\begin{aligned} f(x) = \quad & \sum_m C_m^{\mathrm{B}} X_m + \sum_k C_k^{\mathrm{B}} X_k + \sum_m C_m^{\mathrm{U}} W_m + \sum_k C_k^{\mathrm{U}} D_k \\ h(y) = \quad & \sum_{ij} C_{ij}^{\mathrm{P}} P_{ij} + \sum_{im} C_{im}^{\mathrm{H}} \sum_j Y_{ijm} + \sum_{ik} C_{ik}^{\mathrm{H}} \sum_m Y_{imk} \\ & + \sum_{ijm} C_{ijm}^{\mathrm{T}} Y_{ijm} + \sum_{imk} C_{imk}^{\mathrm{T}} Y_{ijm} + \sum_{ikl} C_{ikl}^{\mathrm{T}} Y_{ikl} \\ X = \quad & \{x | (5.10),\ (5.11) \text{ hold}\} \\ Y(x,\omega) = \quad & \{y | (5.1) \sim (5.9),\ (5.12) \sim (5.16) \text{ hold}\}. \end{aligned}$$

Similarly, we have the mean-MASD model of the form

$$\begin{aligned} \min \quad & f(x) + \mathbb{E}[G(x,\omega)] + \lambda \mathbb{E}[G(x,\omega) - \mathbb{E}G(x,\omega)]_+ \\ \text{s.t.} \quad & x \in X. \end{aligned} \tag{5.18}$$

Both models are two-stage stochastic programs with mixed-integer first-stage and continuous second-stage. The difficulty for solving them is two-fold: on one hand, evaluating $\mathbb{E}[G(x,\omega)]$ or $\delta_+[G(x,\omega)]$ requires solving a large (possibly infinite) number of second-stage network flow problems; on the other hand, the function $\mathbb{E}[G(x,\omega)]$

or $\mathbb{E}[G(x,\omega)]+\lambda\delta_+[G(x,\omega)]$ is nonlinear (typically nonsmooth, although convex) with respect to x. Here we propose a solution strategy by integrating sample average approximation methods (for traditional SP programs and for mean-MAD models) and the branch-and-cut algorithm. SAA converts one stochastic model to a deterministic one, which is then solved by branch-and-cut. For the branch-and-cut algorithm, please see [6] and references therein.

One big concern about stochastic programming models is whether they satisfy the condition of relatively complete recourse. That is, whether all second-stage problems are feasible when the first-stage variables are fixed. For this problem, the second-stage infeasibility of this problem can only come from capacity insufficient to meet demands. To handle this, we include an additional cost in our stochastic models to penalize unmet demands. Define Z_{il} as the quantity of unmet demand on product i at customer zone l, which is nonnegative. Then modify constraint (5.8) as

$$\sum_k Y_{ikl} + Z_{il} \geq D_{il} \quad \forall i,l \tag{5.19}$$

and the objective function of the second-stage problem as

$$\begin{aligned} h(y) = \quad & \sum_{ij} C^{\mathrm{P}}_{ij} P_{ij} + \sum_{im} C^{\mathrm{H}}_{im} \sum_j Y_{ijm} + \sum_{ik} C^{\mathrm{H}}_{ik} \sum_m Y_{imk} \\ & + \sum_{ijm} C^{\mathrm{T}}_{ijm} Y_{ijm} + \sum_{imk} C^{\mathrm{T}}_{imk} Y_{ijm} + \sum_{ikl} C^{\mathrm{T}}_{ikl} Y_{ikl} \\ & + \sum_{il} C^{\mathrm{Z}}_{il} Z_{il}, \end{aligned}$$

where the last term is the cost resulted by shortfall. Except for these two changes, the two stochastic models remain the same. In many situations, the shortfall penalty cost in the above model is a quite subjective parameter, since it is difficult to quantify the loss due to unmet customer demands. In our experiments, we set C^{Z}_{il} equal to the product of some positive constant p_d and the average production cost of that commodity. We call p_d the penalty factor.

With the existence of the infeasibility cost, a simple report of one candidate solution x and the corresponding value $\mathbb{E}[G(x,\omega)]$ does not provide us enough information about the chance of second-stage problem being feasible and actual second-stage cost. Hence we introduce the notion of "probability of infeasibility" and "conditional expectation," where the former $\Pr\{\text{infeas}\}$ represents the probability that a demand scenario can not be satisfied, and the latter $\mathbb{E}[G(x,\omega)|\text{feas}]$ is the expected second-stage cost conditional on feasible demand scenarios. Below we illustrate the principle for estimating $\Pr\{\text{infeas}\}$.

Let $R_1, ..., R_{N_u}$ be an *iid* sequence of N_u Bernoulli trials with probability distribution $\Pr\{R_i = 1\} = p$ and $\Pr\{R_i = 0\} = 1 - p,\ i = 1, \cdots, N_u$. We put $R_i = 0$ if the i-th second-stage problem is feasible, and put $R_i = 1$ otherwise. Then p represents the probability of the second-stage problem in consideration being infeasible, i.e., $\Pr\{\text{infeas}\}$. Let $S \equiv R_1 + ... + R_{N_u}$, i.e., S is the number of times the second-stage problem is infeasible in N_u trials. Suppose $S = k$, where k is a small nonnegative integer (in particular, it may happen that $k = 0$). What can we say about p?

For large N_u and small p, random variable S has Poisson distribution with parameter $\theta = N_u p$. In order for the probability of the event $\{S \leq k\}$,

$$\Pr\{S \leq k\} = \sum_{i=0}^{k} \frac{\theta^i}{i!} e^{-\theta},$$

to be not "small," say at least $\beta > 0$, θ must satisfy

$$\sum_{i=0}^{k} \frac{\theta^i}{i!} e^{-\theta} \geq \beta,$$

which gives an estimate of λ and then $p = \theta / N_u$. This approach is based on confidence intervals. It works only for small k. When k is larger, we can estimate $\Pr\{S \leq k\} \geq \beta$ with the help of Cramér's large deviations theorem. One approximate upper bound is

$$p \leq \frac{k + \sqrt{4k \ln(1/\beta)}}{N_u}.$$

In both cases, we denote by $\Pr\{\text{infeas}\}_\beta$ the p-value corresponding to β.

5.4 Computational Results

In our experiments, the supply chain consists of 2 products, 1 raw material resource, 3 plants, 6 warehouse candidates, 8 distribution center candidates, and 6 customer zones. The two-stage problem has 14 binary variables, 14 continuous variables, and 14 constraints in the first-stage, and 246 continuous variables and 393 constraints in the second-stage. Any production cost C_{ij}^{P} is assumed to be uniformly distributed on interval $\left[0.5\mathbb{E}[C_{ij}^{\mathrm{P}}], 1.5\mathbb{E}[C_{ij}^{\mathrm{P}}]\right]$ with given expected value. The demands D_{il} $\forall i, l$ are assumed to take a multi-variate lognormal distribution, with given expected value vector (μ_{il}). And its covariance matrix is generated from (μ_{il}) and two nonnegative parameters σ_d and ρ_d. The standard deviation of demand D_{il} is $\sigma_d \mu_{il}$, and the correlation coefficient between any two demands is ρ_d.

The SAA parameters we use here are: sample size N for estimating α is 600, sample size N_l for lower bounding is 500, number of samples M_l for lower bounding is 20, sample size N_u for upper bounding is 10000. The best solution reported is the one obtained from lower bounding SAA problems with the best upper bound. The solution algorithm is implemented with C, complied and run on a Linux workstation with dual 2.4 GHz Intel Xeon processors and 2 GB RAM. Here we present the results on the algorithm's performance as well as the SAA solution as functions of weight facotr λ, penalty factor p_d, relative standard deviation σ_d, and correlation coefficient between any two demands ρ_d. Unless otherwise specified, $\lambda = 0.5$, $p_d = 3$, $\sigma_d = 0.5$, and $\rho_d = 0.5$.

Table 5.2 lists notations used for reporting the computational results. Table 5.3 gives a comparison of solutions from three models, where $\lambda = 0.5$ in the mean-MASD model. Tables 5.4, 5.5, 5.6, and 5.7 report computational results of the mean-MASD model as a function of λ, σ_d, ρ_d, and p_d, respectively. Moreover, we provide the plots of UB, LB, and $\Pr\{\text{infeas}\}_{0.95}$ with respect to the four parameters in Figures 5.2 and 5.3.

Table 5.2: Notation for supply chain network design problem

p_d	penalty factor ($p = P\times$ production cost of the demand)
σ_d	relative stand. dev. (the ratio of one demand's stand. dev. to its mean)
ρ_d	correlation coefficient between two demands
$\bar{\alpha}$	the estimated α-value from Step 1 of SAA for mean-MASD models
UB	upper bound on the optimal mean-MASD value given by SAA
UBSD	stand. dev. of UB
LB	lower bound on the optimal mean-MASD value given by SAA
LBSD	stand. dev. of LB
UB-mean	the mean part of UB
UB-MASD	the MASD part of UB
cost_1	the first-stage cost corresponding to the solution $\hat{x}$ providing UB
$\hat{\mathbb{E}}[\text{cost}_2]$	the sample average of second-stage cost
$\hat{\mathbb{E}}[\text{cost}_2]$ SD	stand. dev. of $\hat{\mathbb{E}}[\text{cost}_2]$
$\Pr\{\text{infeas}\}_{0.95}(\%)$	the 95%-confident probability that the second-stage problem is infeasible, resulted by $\hat{x}$
$\hat{\mathbb{E}}[\text{cost}_2\|\text{feas}]$	the estimated expected second-stage cost conditional on feasible second-stage problem total-time total time consumed (in seconds)
total-time	time in seconds spent on the solution of one model
W1 (1300)	the actual capacity of the 1st warehouse candidate in $\hat{x}$ with maximal capacity being 1300, so are other warehouse candidates
D1 (1000)	the actual capacity of the 1st distribution center candidate in $\hat{x}$ with maximal capacity being 1000, so are other warehouse candidates

5.4.1 Comparison of Three Models

Table 5.3 gives a comparison of solutions from three models. The second column corresponds to the solution from the deterministic model with production costs and demands taking the expected values of their counterparts in stochastic models. The third and fourth column correspond to one randomly chosen solution from the lower bounding problem of the traditional SP model and the mean-MASD model, respectively. We use a sample of size N_u for evaluating the quality of these solutions. Table 5.3 clearly shows the successive improvement in the robustness of the solution as one adopts the deterministic, traditional SP, and Mean-MASD models. Specifically, going from left to right, more warehouses and distribution centers enter the optimal solutions. This results in an increase in both the first-stage cost and the conditional second-stage cost. We also notice that the deterministic model is solved much faster than the stochastic models.

5.4.2 Effects of λ

In Table 5.4, as λ goes from 0.1 to 0.9, UB, LB, UB-MASD increase, and $\Pr\{\text{infeas}\}_{0.95}$ decreases. Since this is a mixed-integer problem, the changes are not continuous, e.g. these values remain the same when λ goes from 0 to 0.3.

5.4.3 Effects of σ_d and ρ_d

Tables 5.5 and 5.6 show that as σ_d and ρ_d increase, the upper and lower bound increase, while $\Pr\{\text{infeas}\}_{0.95}$ fluctuates and shows no tendency in the range we study. This means that, as σ_d or ρ_d increases, random demands have more volatility; to compensate this increased volatility, we need to spend more money to control the risk and get roughly same level of infeasibility probability of second-stage problems.

5.4.4 Effects of p_d

As we pointed before, p_d is a parameter difficult to measure. In Table 5.7, as p_d increases from 2 to 9, i.e., we emphasize more on getting feasible second-stage problems, UB, LB, UB-mean increase, UB-MASD fluctuates, and $\Pr\{\text{infeas}\}_{0.95}$ decreases.

Table 5.3: Comparison of three models (Supply Chain)

Model	Deterministic	Traditional SP	Mean-MASD
cost_1	17541	22095	26556
$\hat{\mathbb{E}}[\text{cost}_2]$	144834	130156	126408
$\hat{\mathbb{E}}[\text{cost}_2]$ SD	861	681	600
$\Pr\{\text{infeas}\}_{0.95}(\%)$	43.52	13.29	4.39
$\hat{\mathbb{E}}[\text{cost}_2 \vert \text{feas}]$	109076	110104	118032
total-time	7.57	1414.32	1694.10
W1 (1300)	0	0	620
W2 (1100)	0	0	0
W3 (1200)	0	0	0
W4 (1200)	0	0	0
W5 (1100)	1100	1100	1100
W6 (1300)	610	1300	1239
D1 (1000)	800	1000	1000
D2 (900)	0	0	0
D3 (950)	0	0	0
D4 (1050)	910	1050	1050
D5 (1100)	0	0	0
D6 (1000)	0	0	0
D7 (980)	0	350	909
D8 (1050)	0	0	0

Table 5.4: Effects of λ (Supply Chain, mean-MASD)

λ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\bar{\alpha}$	0.633	0.633	0.633	0.610	0.608	0.608	0.608	0.610	0.610	0.608
UB	152251	154561	156872	159182	161354	163444	165541	167637	169732	171825
UBSD	681	664	648	633	549	524	510	497	485	473
LB	150607	152892	155087	157319	159430	161570	163664	165765	167774	169811
LBSD	758	767	776	789	792	792	807	812	834	826
UB-mean	152251	152251	152251	152251	152876	152964	152965	152968	152977	152989
UB-MASD	23104	23104	23104	23104	21195	20961	20959	20955	20944	20928
cost_1	22095	22095	22095	22095	26077	26556	26559	26570	26597	26635
$\Pr\{\text{infeas}\}_{0.95}(\%)$	13.29	13.29	13.29	13.29	5.60	4.39	4.38	4.32	4.24	4.18
$\hat{\mathbb{E}}[\text{cost}_2 \vert \text{feas}]$	110104	110104	110104	110104	116721	118032	118042	118103	118208	118275
total-time	1684	1589	1536	1543	1525	1694	1673	1692	1665	1641
W1 (1300)	0	0	0	0	642	620	621	618	628	634
W2 (1100)	0	0	0	0	0	0	0	0	0	0
W3 (1200)	0	0	0	0	0	0	0	0	0	0
W4 (1200)	0	0	0	0	0	0	0	0	0	0
W5 (1100)	1100	1100	1100	1100	1100	1100	1100	1100	1100	1100
W6 (1300)	1300	1300	1300	1300	1108	1239	1239	1244	1241	1243
D1 (1000)	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
D2 (900)	0	0	0	0	0	0	0	0	0	0
D3 (950)	0	0	0	0	0	0	0	0	0	0
D4 (1050)	1050	1050	1050	1050	1050	1050	1050	1050	1050	1050
D5 (1100)	0	0	0	0	0	0	0	0	0	0
D6 (1000)	0	0	0	0	0	0	0	0	0	0
D7 (980)	350	350	350	350	800	909	910	912	919	927
D8 (1050)	0	0	0	0	0	0	0	0	0	0

Table 5.5: Effects of σ_d (Supply Chain, mean-MASD)

σ_d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\bar{\alpha}$	0.542	0.548	0.565	0.597	0.608	0.627	0.652	0.665	0.678
UB	144847	148291	153284	158087	163444	168097	173259	178716	183775
UBSD	136	221	303	443	524	675	850	1036	1136
LB	144662	147722	152393	156701	161585	165921	170438	175078	179629
LBSD	236	360	481	656	795	964	1204	1397	1599
UB-mean	141433	143232	146599	149086	152964	155422	158214	161252	165213
UB-MASD	6828	10118	13370	18001	20961	25350	30089	34928	37125
cost_1	18555	18955	22095	22095	26556	26867	26867	26867	31882
$\Pr\{\text{infeas}\}_{0.95}(\%)$	3.53	9.89	4.94	9.73	4.39	6.25	8.22	9.76	6.12
$\hat{\mathbb{E}}[\text{cost}_2 \vert \text{feas}]$	121854	118928	119699	115077	118032	114653	110432	106388	110135
total-time	1237	1206	1318	1452	1605	1545	1374	1318	1238
W1 (1300)	0	0	0	0	620	681	700	705	1300
W2 (1100)	0	0	0	0	0	0	0	0	0
W3 (1200)	0	0	0	0	0	0	0	0	0
W4 (1200)	0	0	0	0	0	0	0	0	0
W5 (1100)	1100	1100	1100	1100	1100	1100	1100	1100	1100
W6 (1300)	853	950	1300	1300	1239	1249	1230	1225	1300
D1 (1000)	980	1000	1000	1000	1000	1000	1000	1000	1000
D2 (900)	0	0	0	0	0	0	0	0	0
D3 (950)	0	0	0	0	0	0	0	0	670
D4 (1050)	973	1050	1050	1050	1050	1050	1050	1050	1050
D5 (1100)	0	0	0	0	0	0	0	0	0
D6 (1000)	0	0	0	0	0	0	0	0	0
D7 (980)	0	0	350	350	909	980	980	980	980
D8 (1050)	0	0	0	0	0	0	0	0	0

Table 5.6: Effects of ρ_d (Supply Chain, mean-MASD)

ρ_d	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\bar{\alpha}$	0.558	0.580	0.582	0.598	0.618	0.608	0.620	0.618	0.630	0.627
UB	149075	153137	155905	158658	161300	163444	165192	166858	168448	169951
UBSD	232	291	376	457	531	524	578	631	682	729
LB	148788	152555	155050	157377	159661	161585	163184	164769	165970	167465
LBSD	306	385	495	605	723	795	904	1052	1176	1328
UB-mean	143794	146593	147950	149402	150890	152964	153902	154789	155628	156425
UB-MASD	10561	13088	15909	18511	20821	20961	22581	24138	25640	27053
cost_1	18955	22095	22095	22095	22530	26556	26755	26867	26867	726867
$\Pr\{\text{infeas}\}_{0.95}(\%)$	10.36	4.64	7.76	10.13	11.84	4.39	4.93	5.58	6.21	6.98
$\hat{\mathbb{E}}[\text{cost}_2 \vert \text{feas}]$	118876	120118	117284	114647	111964	118032	117038	115844	114650	113277
total-time	1267	1386	1500	1530	1583	1632	1481	1512	1446	1446
W1 (1300)	0	0	0	0	0	620	617	668	680	678
W2 (1100)	0	0	0	0	0	0	0	0	0	0
W3 (1200)	0	0	0	0	0	0	0	0	0	0
W4 (1200)	0	0	0	0	0	0	0	0	0	0
W5 (1100)	1100	1100	1100	1100	1100	1100	1100	1100	1100	1100
W6 (1300)	950	1300	1300	1300	1300	1239	1288	1262	1250	1252
D1 (1000)	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
D2 (900)	0	0	0	0	0	0	0	0	0	0
D3 (950)	0	0	0	0	350	0	0	0	0	0
D4 (1050)	1050	1050	1050	1050	1050	1050	1050	1050	1050	1050
D5 (1100)	0	0	0	0	0	0	0	0	0	0
D6 (1000)	0	0	0	0	0	0	0	0	0	0
D7 (980)	0	350	350	350	0	909	954	980	980	980
D8 (1050)	0	0	0	0	0	0	0	0	0	0

Table 5.7: Effects of p_d (Supply Chain, mean-MASD)

p_d	2	3	4	5	6	7	8	9
$\bar{\alpha}$	0.612	0.608	0.610	0.617	0.622	0.628	0.608	0.608
UB	157560	163444	164949	166407	167701	168189	168585	168972
UBSD	505	524	575	637	565	596	603	633
LB	155954	161585	163122	164171	165235	165805	166496	166575
LBSD	724	795	808	890	889	879	847	842
UB-mean	146775	152964	154190	155304	157262	157645	158069	158365
UB-MASD	21569	20961	21519	22206	20878	21088	21034	21214
cost_1	18955	26556	26867	26867	31330	31368	31882	31882
$\Pr\{\text{infeas}\}_{0.95}(\%)$	25.95	4.39	3.86	3.86	1.49	1.46	1.07	1.07
$\mathbb{E}[\text{cost}_2 \vert \text{feas}]$	101550	118032	118616	118616	121341	121392	122062	122062
total-time	1362	1601	1445	1485	1573	1379	1371	1354
W1 (1300)	0	620	650	650	1177	1186	1300	1300
W2 (1100)	0	0	0	0	0	0	0	0
W3 (1200)	0	0	0	0	0	0	0	0
W4 (1200)	0	0	0	0	0	0	0	0
W5 (1100)	1100	1100	1100	1100	1100	1100	1100	1100
W6 (1300)	950	1239	1280	1280	1300	1300	1300	1300
D1 (1000)	1000	1000	1000	1000	1000	1000	1000	1000
D2 (900)	0	0	0	0	0	0	0	0
D3 (950)	0	0	0	0	547	556	670	670
D4 (1050)	1050	1050	1050	1050	1050	1050	1050	1050
D5 (1100)	0	0	0	0	0	0	0	0
D6 (1000)	0	0	0	0	0	0	0	0
D7 (980)	0	909	980	980	980	980	980	980
D8 (1050)	0	0	0	0	0	0	0	0

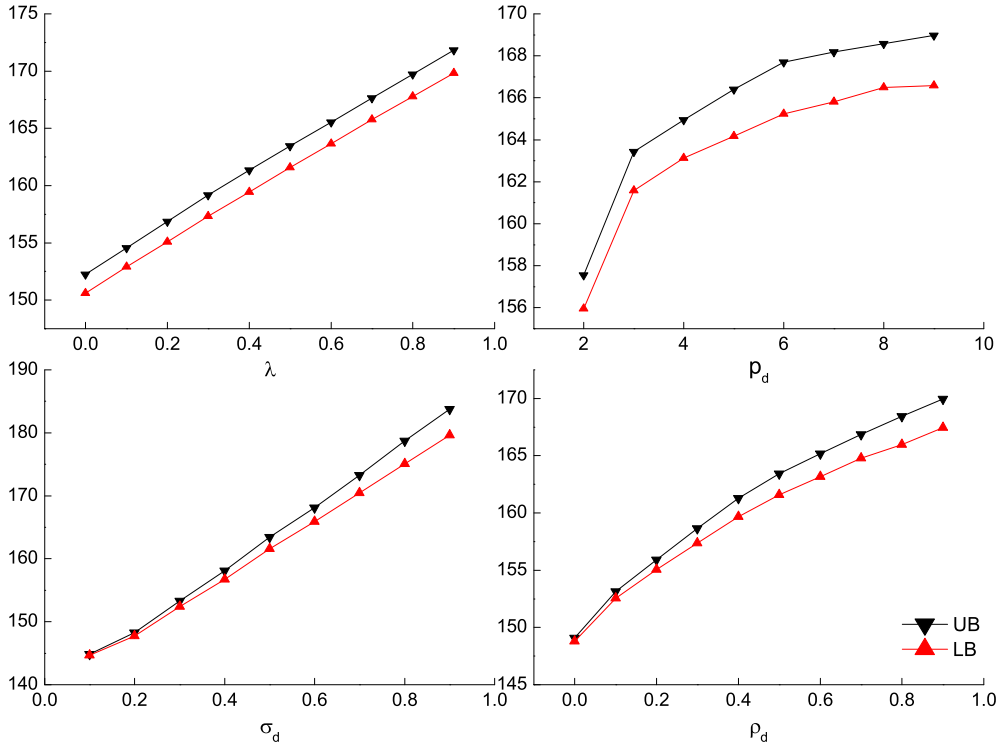


Figure 5.2: UB and LB as functions of λ, p_d, σ_d, and ρ_d (mean-MASD, unit 1000)

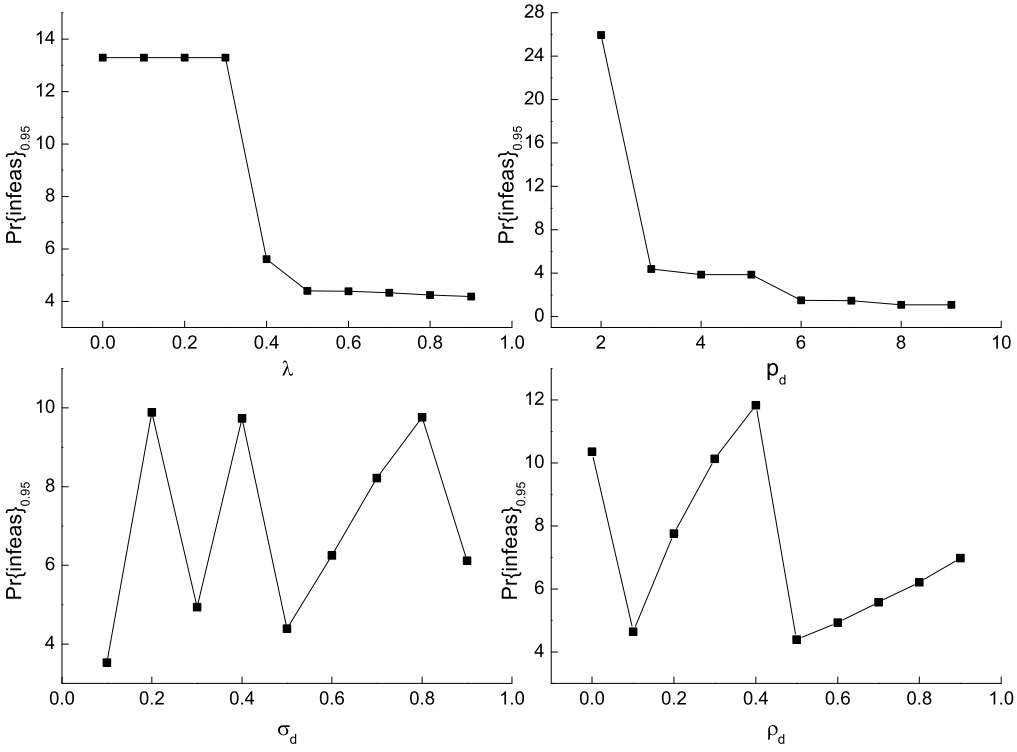


Figure 5.3: $\Pr\{\text{infeas}\}_{0.95}$ as function of λ, p_d, σ_d, and ρ_d (mean-MASD)

CHAPTER VI

CONCLUSIONS

In this thesis we established sample average approximation methods for two classes of non-traditional stochastic programs: stochastic min-max programs and expected value constrained programs, and customized the SAA methods for mean-MAD models and CVaR constrained problems, respectively. In summary, we have made the following specific contributions.

1. We derived the exponential rates at which the optimal value and an optimal solution of the SAA problem converge to their true counterparts with probability approaching one as the sample size increases.

2. We designed practical schemes to implement the SAA methods. Our schemes can return one or more candidate solutions, statistical upper and lower bound estimates on the true optimal value, as well as confidence intervals or probability associated with these bounds. The lower bounding schemes are based on appropriate problem relaxations.

3. We carried out computational studies on a portfolio selection problem and a supply chain network design problem. We formulated both problems as mixed-integer programs. We found that the overall performance of the proposed SAA methods is satisfactory. In the portfolio selection problem, besides testing SAA, we explored and compared different types of risk-averse models (mean-MASD, mean-QDEV, mean-CVaR, MAD, variance, chance constrained, and CVaR constrained), sampling techniques (simple Monte Carlo and Latin Hypercube sampling), probability distributions (uniform discrete, multivariate normal and lognormal), and time-division approaches. In the supply chain problem, we showed the progressive improvement on solution

robustness as one employs more and more risk-sensitive models, i.e., going from deterministic, to traditional SP, then to mean-MASD models. We also investigated the impact of several parameters, including demands variability and penalty cost on unmet demands, in the framework of mean-MASD.

Finally we point out some directions for future research. First, regarding the approach for obtaining exponential convergence rates of mean-MAD models, instead of following the route of specializing the general results of stochastic min-max programs, one may start directly from mean-MAD models by utilizing properties of statistical estimates of MAD [29]. Second, the convergence results we obtained are under *iid* sampling. In cases where variance reduction techniques are used, these convergence rates may not hold. For traditional SP, convergence rates under non-*iid* sampling were discussed in [10, 20]. Similar studies on non-traditional SP will be interesting and useful. Third, besides Latin Hypercube sampling, one may also try other variance reduction techniques, e.g. Quasi-Monte Carlo [23]. Fourth, the SAA schemes we designed involve solving multiple SAA problems. It is worthwhile to design schemes which only needs to solve one or two SAA problems, like those for traditional SP [5].

REFERENCES

[1] S.Ahmed and A.Shapiro. The sample average approximation method for stochastic programs with integer recourse. Available online *http://www.isye.gatech.edu/∼sahmed*, 2002 (date accessed: August 2007).

[2] S.Ahmed. Convexity and decomposition of mean-risk stochastic programs. *Mathematical Progamming*, 106:433-446, 2006.

[3] P.Artzner et al. Coherent measures of risk. *Mathematical Finance*, 9:203-228, 1999.

[4] J.Atlason, M.Epelman and G.Henderson. Call center staffing with simulation and cutting plane methods. *Annals of Operations Research*, 127:333-358, 2004.

[5] G.Bayraksan and D.P.Morton. Assessing solution quality in stochastic programs. *Mathematical Programming*, 108:495-514, 2007.

[6] J.Beasley. *Advances in Linear and Integer Programming*, Oxford University Press, Oxford, 1996.

[7] S.Benati. The optimal portfolio problem with coherent risk measure constraints. *European Journal of Operational Research*, 150:572-584, 2003.

[8] F.Black and M.Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637-654, 1973.

[9] D.Dentcheva and A.Ruszczyński. Portfolio optimization with stochastic dominance constraints. *Journal of Banking and Finance*, 30:433-451, 2006.

[10] S.Drew and T.Homem-de-Mello. Some large deviations results for Latin Hypercube sampling. Available online *http://users.iems.northwestern.edu/∼tito/list.htm*, 2006 (date accessed: August 2007).

[11] J.Dupačová. Uncertainties in stochastic progrogramming models: the minimax approach. Available online *http://drops.dagstuhl.de/opus/volltexte/2005/45*, 2005 (date accessed: August 2007).

[12] A.J.Kleywegt et al. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12:479-502, 2001.

[13] H.Konno and T.Koshizuka. Mean-absolute deviation model. *IIE Transactions*, 37:893-900, 2005.

[14] P.Krokhmal, J.Palmquist and S.Uryasev. Portfolio optimization with conditional value-at-risk objective and constraints. *The Journal of Risk*, 4:11-27, 2002.

[15] D.Kuhn. Convergent bounds for stochastic programs with expected value constraints. Available online *http://edoc.hu-berlin.de/series/speps/2006-22/PDF/22.pdf*, 2006 (date accessed: August 2007).

[16] J.Linderoth, A.Shapiro and S.Wright. The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research*, 142:215-241, 2006.

[17] H.M.Markowitz. Portfolio selection. *Journal of Finance*, 7:77-91, 1952.

[18] H.M.Markowitz. *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, Blackwell Publishing, Oxford, UK, 1987.

[19] M.D.McKay, R.J.Beckman and W.J.Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42:55-61, 2000.

[20] T.Homem-de-Mello. On rates of convergence for stochastic optimization problems under non-I.I.D. sampling. Available online *http://www.optimization-online.org*, 2006 (date accessed: August 2007).

[21] A.Nemirovski and A.Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17:969-996, 2006.

[22] J.V.Neumann and O.Morgenstern. *Theory of Games and Economic Behavior*, Princeton University Press, 1944.

[23] H.Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.

[24] M.O'Brien. *Techniques for Incorporating Expected Value Constraints into Stochastic Programs*, PhD Thesis, Standard University, 2000.

[25] W.Ogryczak and A.Ruszczyński. From stochastic dominance to mean-risk models: semideviations as risk measures. *European Journal of Operational Research*, 116:33-50, 1999.

[26] W.Ogryczak and A.Ruszczyński. On consistency of stochastic dominance and mean-semideviation models. *Mathematical Programming*, 89:217-232, 2001.

[27] W.Ogryczak and A.Ruszczyński. Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization*, 13:60-78, 2002.

[28] G.Ch.Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. In: S.Uryasev (ed.), *Probabilistic Constrained Optimization: Methodology and Applications*, Kluwer, Boston, 2000.

[29] T.Pham-Gia and T.L.Hung. The mean and median absolute deviations. *Mathematical and Computer Modelling*, 34:921-936, 2001.

[30] A.Prákopa. *Stochastic Programming*, Kluwer, Boston, 1995.

[31] R.T.Rockafellar and S.Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2:21-41, 2000.

[32] R.T.Rockafellar and S.Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26:1443-1471, 2002.

[33] A.Ruszczyński and R.J.Vanderbei. Frontiers of stochastically nondominated portfolios. *Econometrica*, 71:1287-1297, 2003.

[34] A.Ruszczyński and A.Shapiro (eds.), *Stochastic Programming,* in: Handbooks in OR & MS, Vol.10, Elsevier Science, Amsterdam, 2003.

[35] A.Ruszczyński and A.Shapiro. Optimization of risk functions. In: G.Calafiore and F.Dabbene (eds.), *Probabilistic and Randomized Methods for Design under Uncertainty*, Springer-Verlag, London, 2006.

[36] T.Santoso, S.Ahmed, M.Goetschalckx and A.Shapiro. A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research*, 167:96-115, 2005.

[37] R.Schultz. Stochastic programming with integer variables. *Mathematical Programming*, 97:285-309, 2003.

[38] A.Shapiro and A.J.Kleywegt. Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17:523-542, 2002.

[39] A.Shapiro and S.Ahmed. On a class of minimax stochastic programs. *SIAM Journal on Optimization*, 14:1237-1249, 2004.

[40] A.Shapiro and H.Xu. Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation. Available online *http://www.optimization-online.org*, 2005 (date accessed: August 2007).

[41] A.Shapiro. Asymptotics of minimax stochastic programs. Available online *http://www.optimization-online.org*, 2006 (date accessed: August 2007).

[42] A.Shapiro. Stochastic programming approach to optimization under uncertainty. *Mathematical Programming*, 112:183-220, 2008.

[43] L.V.Snyder. Facility location under uncertainty: a review. *IIE Transactions*, 38:537-554, 2006.

[44] D.W.Stroock. *An Introduction to the Theory of Large Deviations*, Springer-Verlag, New York, 1984.

[45] P.Tsiakis, N.Shah, and C.C.Pantelides. Design of multi-echelon supply chain networks under demand uncertainty. *Industrial and Engineering Chemistry Research*, 40:3585-3604, 2001.